Geospatial Visual Analytics:
Visual analysis of spatio-temporal data
Part 3: behaviours of moving objects

Fraunhofer Institut
Intelligente Analyse- und Informationssysteme
IAIS

Gennady Andrienko & Natalia Andrienko
http://geoanalytics.net

Types of Temporal Variance

- Changes of thematic properties (values of attributes) associated with places
  - e.g. district population, data from stationary sensors
- Existential changes (appearance and disappearance)

  Events: objects with limited life time
  - e.g. earthquakes, traffic incidents, observations of rare plants or animals
- Changes of spatial properties: location, size, shape, orientation, altitude, etc.
  - e.g. movement of vehicles, growth of cities

Types of Temporal Variance

- Changes of thematic properties (values of attributes) associated with places
  - e.g. district population, data from stationary sensors
- Existential changes (appearance and disappearance)

  Events: objects with limited life time
  - e.g. earthquakes, traffic incidents, observations of rare plants or animals
- Changes of spatial properties: <u>**location**</u>, size, shape, orientation, altitude, etc.
  - e.g. <u>**movement**</u> of vehicles, growth of cities

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Example: movement of a person using his private car

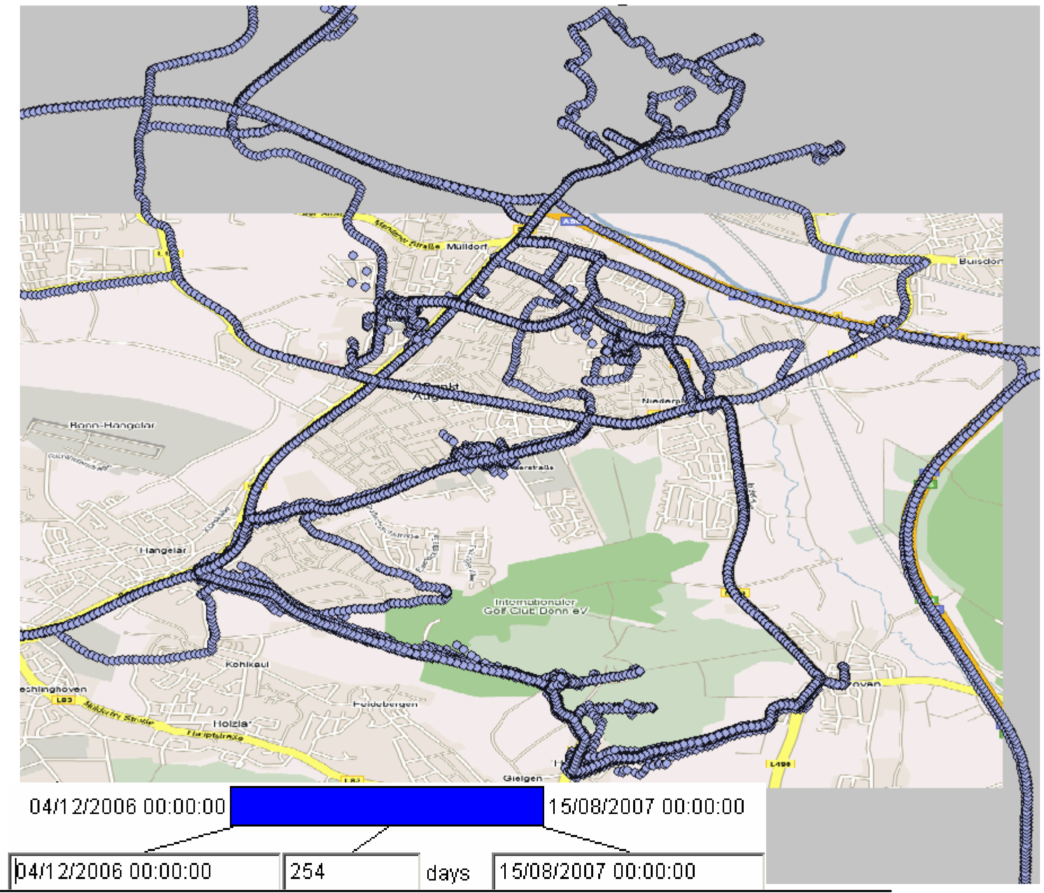
The data come from a GPS tracking device

N;Time;Lat;Long;Height;Course;Speed;PDOP;State;NSat
…
8;22/03/07 08:51:52;50.777132;7.205580; 67.6;345.4;21.817;3.8;1808;4
9;22/03/07 08:51:56;50.777352;7.205435; 68.4;35.6;14.223;3.8;1808;4
10;22/03/07 08:51:59;50.777415;7.205543; 68.3;112.7;25.298;3.8;1808;4
11;22/03/07 08:52:03;50.777317;7.205877; 68.8;119.8;32.447;3.8;1808;4
12;22/03/07 08:52:06;50.777185;7.206202; 68.1;124.1;30.058;3.8;1808;4
13;22/03/07 08:52:09;50.777057;7.206522; 67.9;117.7;34.003;3.8;1808;4
14;22/03/07 08:52:12;50.776925;7.206858; 66.9;117.5;37.151;3.8;1808;4
15;22/03/07 08:52:15;50.776813;7.207263; 67.0;99.2;39.188;3.8;1808;4
16;22/03/07 08:52:18;50.776780;7.207745; 68.8;90.6;41.170;3.8;1808;4
17;22/03/07 08:52:21;50.776803;7.208262; 71.1;82.0;35.058;3.8;1808;4
18;22/03/07 08:52:24;50.776832;7.208682; 68.6;117.1;11.371;3.8;1808;4
…

About 100,000 positions

Data structure:

<time, position, {any other attributes}>



Analysis of movements of an individual *(person, animal, etc.)***: major tasks and typical problems**

- Investigate the movement behaviour of the object:
  - Detect and interpret typical trips: sources, destinations, routes, intermediate stops, purposes, …
  - Find out the typical times of the trips
  - If possible, interpret also atypical trips (e.g. unusual route, timing, duration, …)
- <u>Problems</u>:
  - no explicit trips in the data but just positions
  - no meaningful places but coordinates

⇒ Places and trips have to be defined through data analysis!

✓ *In our example, we know that the positions were recorded only while the car moved. Otherwise, the data could be filtered to remove sequences of records made in the same place (using an appropriate distance threshold).*

- <u>Problem</u>: the position records are too numerous to be examined by a human analyst

⇒ The use of database processing and computational techniques is strictly necessary!

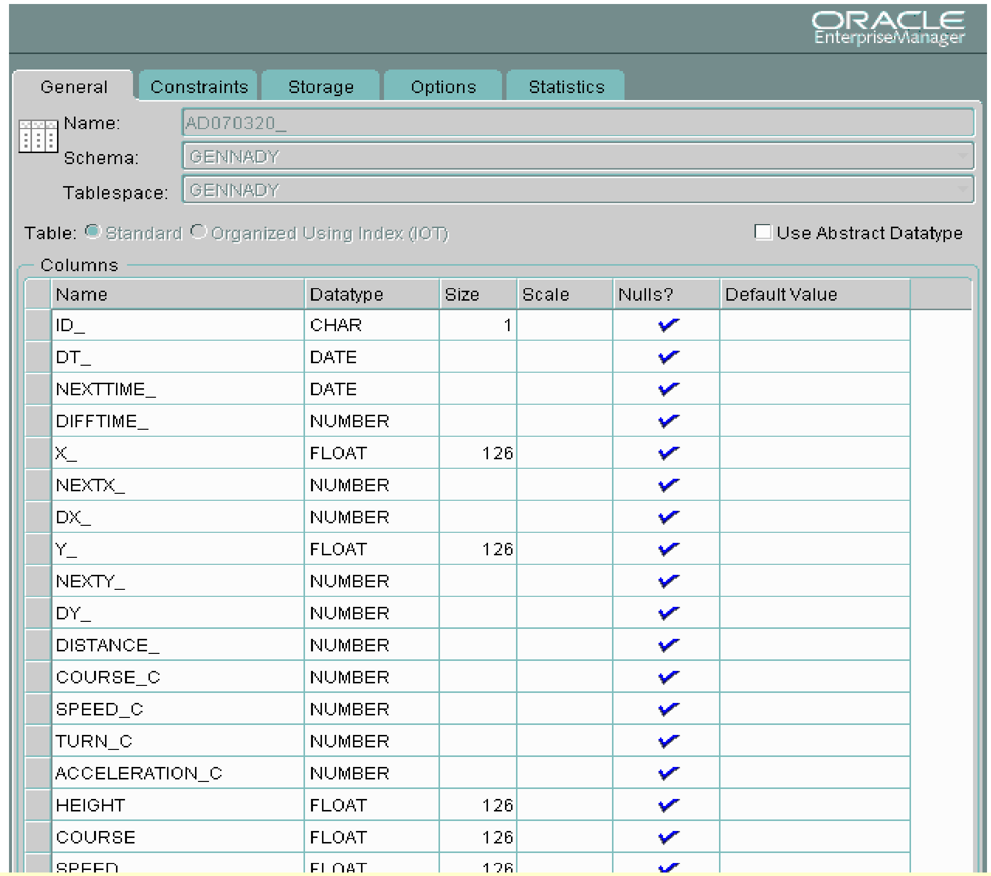
Data preprocessing (one-time operation in DB)

Original data:

- <t, x, y>

Results of preprocessing:

- Points are arranged into temporally ordered sequences (NEXTTIME field)
- Computed distances in space and time to the next position (DIFFTIME and DISTANCE)
- Derived speed, course, acceleration and turn in each point
- Temporal components are easy to extract: day of week, day of year, decade of month…



Name	Datatype	Size	Scale	Nulls?	Default Value
ID_	CHAR	1		✓	
DT_	DATE			✓	
NEXTTIME_	DATE			✓	
DIFFTIME_	NUMBER			✓	
X_	FLOAT	126		✓	
NEXTX_	NUMBER			✓	
DX_	NUMBER			✓	
Y_	FLOAT	126		✓	
NEXTY_	NUMBER			✓	
DY_	NUMBER			✓	
DISTANCE_	NUMBER			✓	
COURSE_C	NUMBER			✓	
SPEED_C	NUMBER			✓	
TURN_C	NUMBER			✓	
ACCELERATION_C	NUMBER			✓	
HEIGHT	FLOAT	126		✓	
COURSE	FLOAT	126		✓	
SPEED	FLOAT	126		✓	

This time-consuming data pre-processing needs to be performed only once.
It is therefore separated from further analysis and aggregation procedures.

Subtask 1: define significant places

- We know that the DB contains position records only for the times when the car moved.

⇒ Large time gaps between successive positions indicate stops.

⇒ We can query the database for the positions where DIFFTIME (distance in time to the next position) exceeds a specified threshold

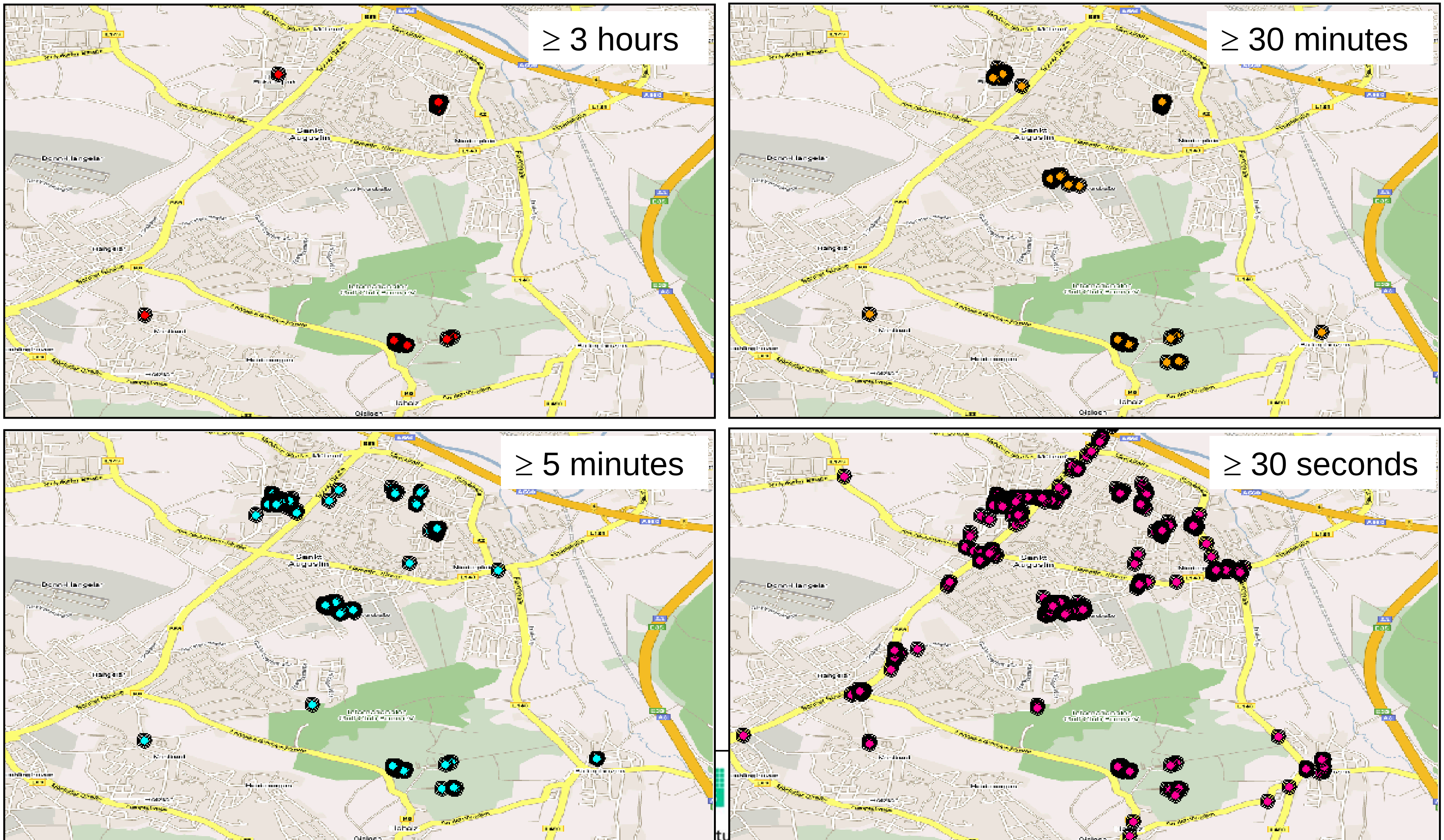
- The possible meanings of the extracted places will differ depending on the threshold
  - Large gaps: places where the person spends much time (home, work)
  - Medium gaps: shops, doctors, sport facilities, …
  - Small gaps: traffic lights, street crossings, …
- Places separated by medium to large gaps may be destinations and sources of trips

  ⇒ Sequences of positions between the gaps represent the trips

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Extraction of the places of stops

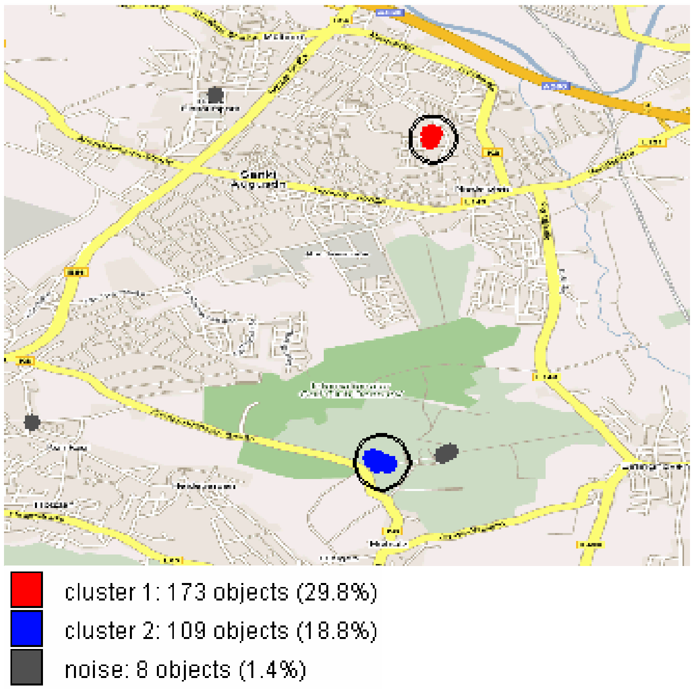
DB query: select points where DIFTIME >= chosen temporal threshold



Interpretation of the places of stops
A) Long stops (≥ 3 hours)

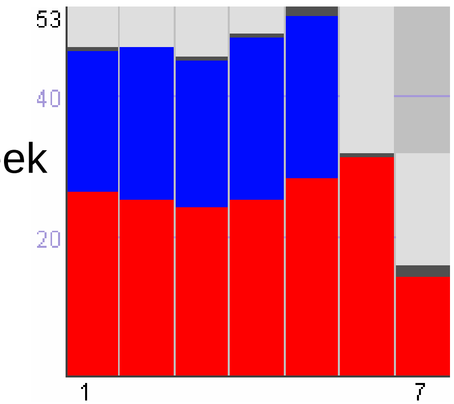
1) Spatial clustering: find places of <u>repeated</u> stops

(more significant than occasional stops)

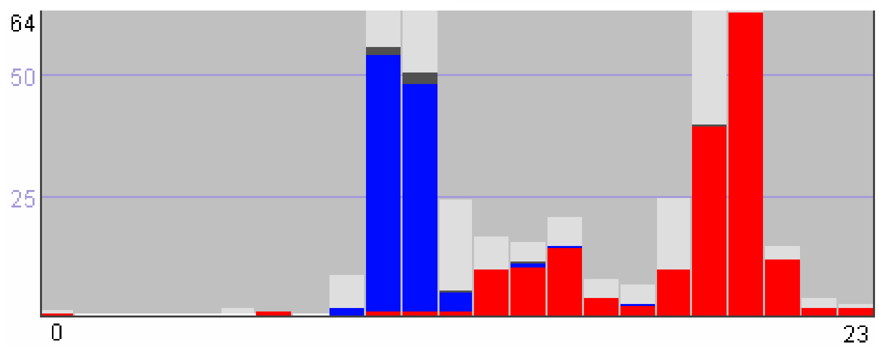


2) Look at the days and times of the occurrence

Days of the week



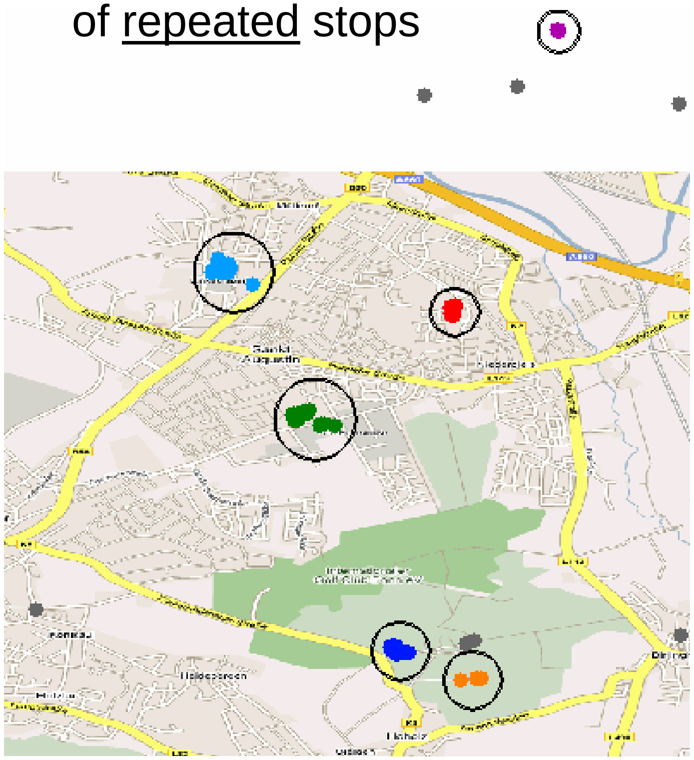
Hours of the day



⇒ Red: home, blue: work

Interpretation of the places of stops
B) Medium stops (≥ 30 minutes)

1) Spatial clustering: find places of <u>repeated</u> stops

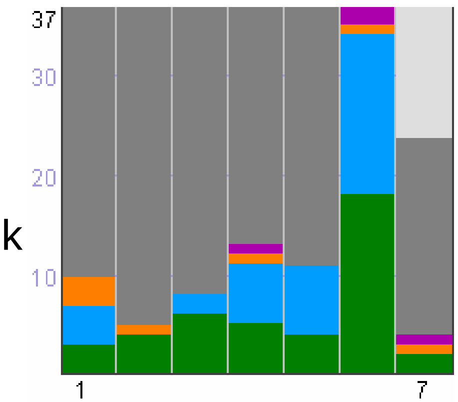


- cluster 1: 184 objects (22.7%)
- cluster 2: 117 objects (14.4%)
- cluster 3: 42 objects (5.2%)
- cluster 4: 35 objects (4.3%)
- cluster 5: 7 objects (0.9%)
- cluster 6: 4 objects (0.5%)
- noise: 17 objects (2.1%)

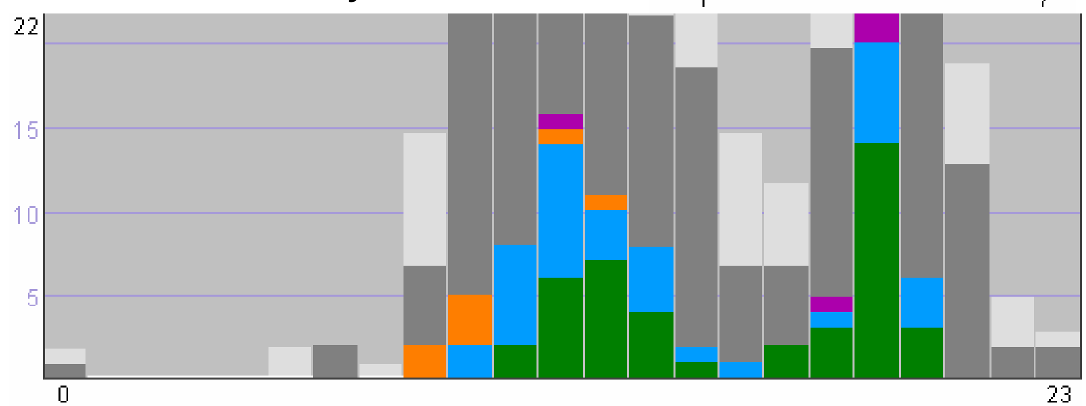
2) Look at the days and times of the occurrence

(clusters 1&2 excluded)

Days of the week



Hours of the day

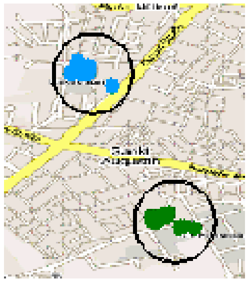


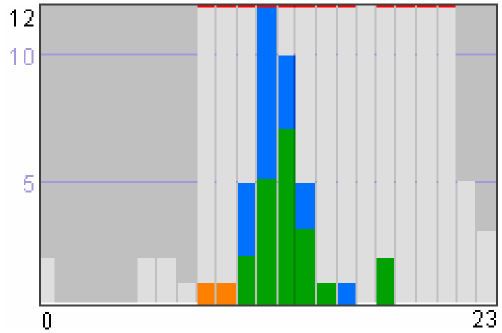
⇒ Green & light blue: probably shopping

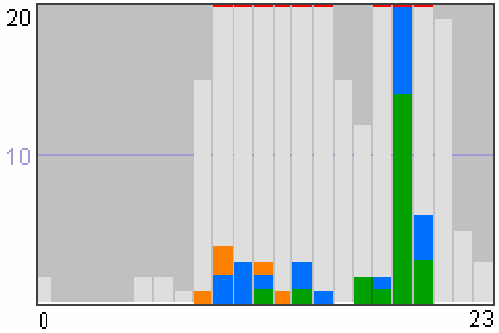
Interpretation of the clusters 3 and 4 (continued)

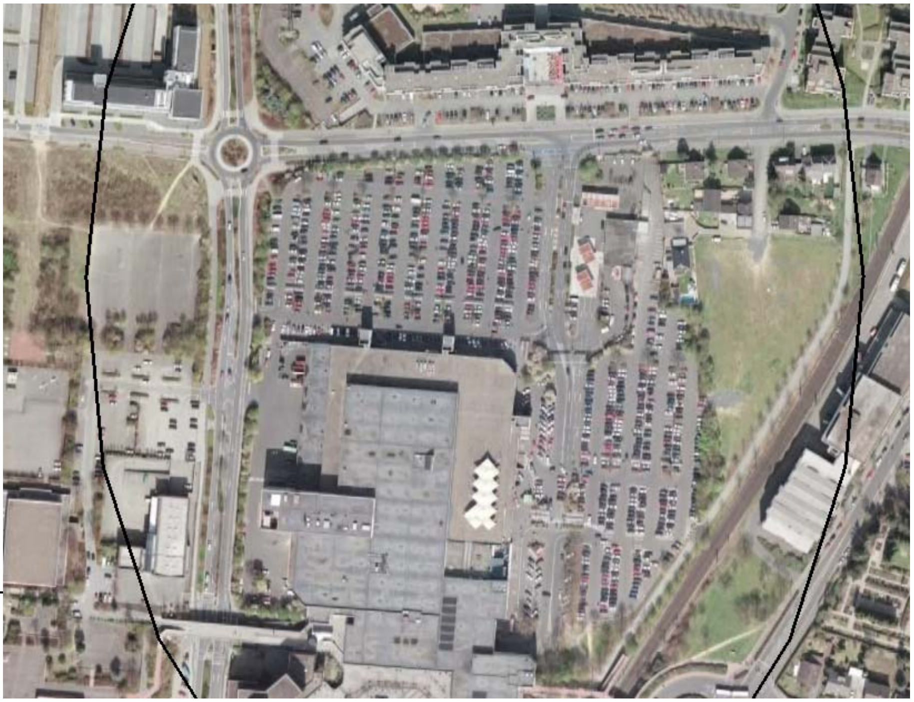
Saturday and Sunday: the stops mostly occur between 10 and 14

Monday to Friday: the stops mostly occur in the evening hours (max between 18 and 19)

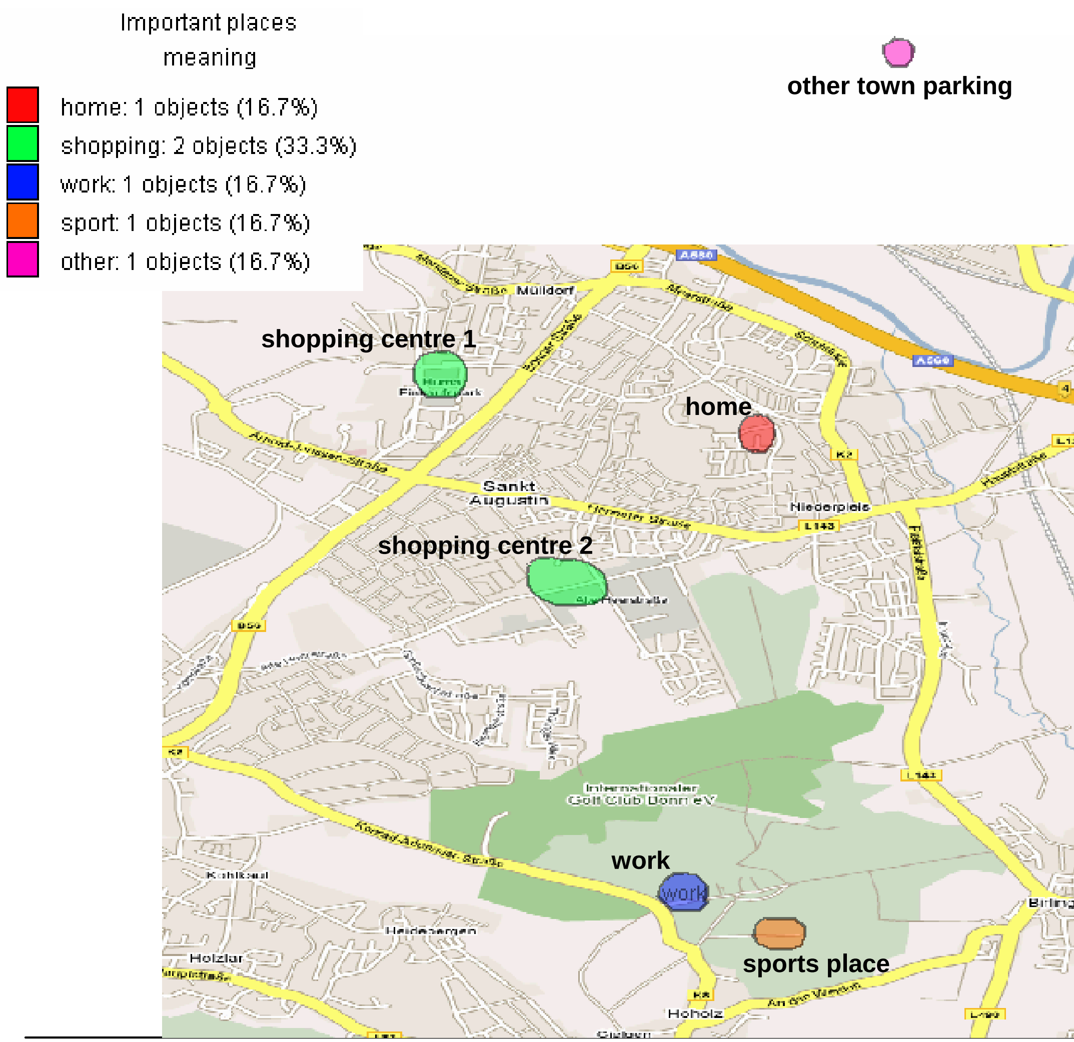


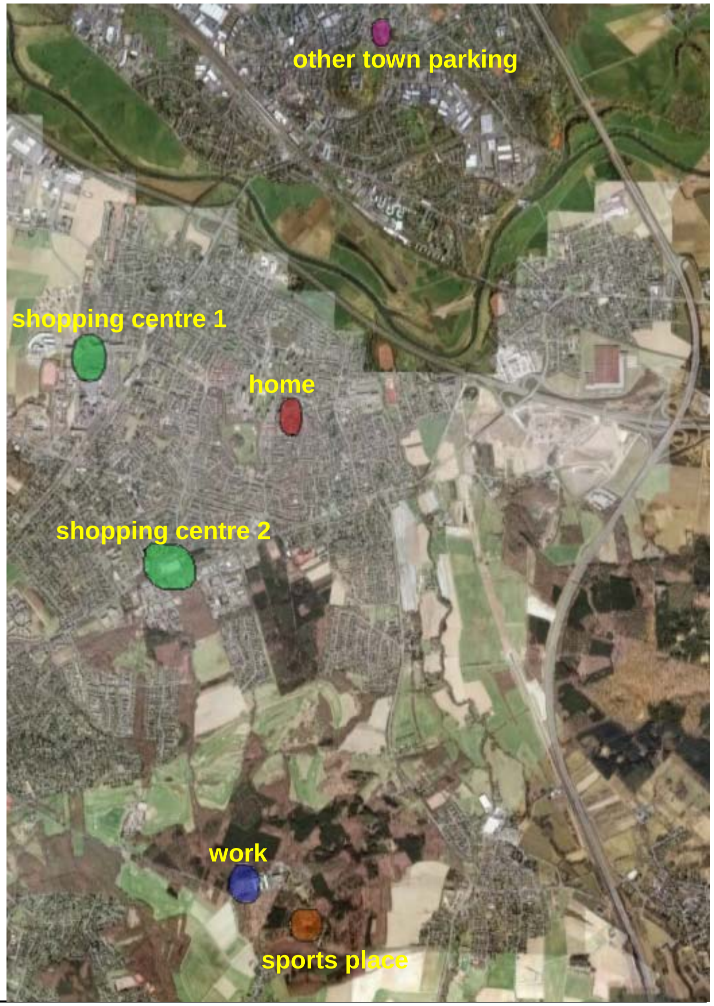






The interpreted places

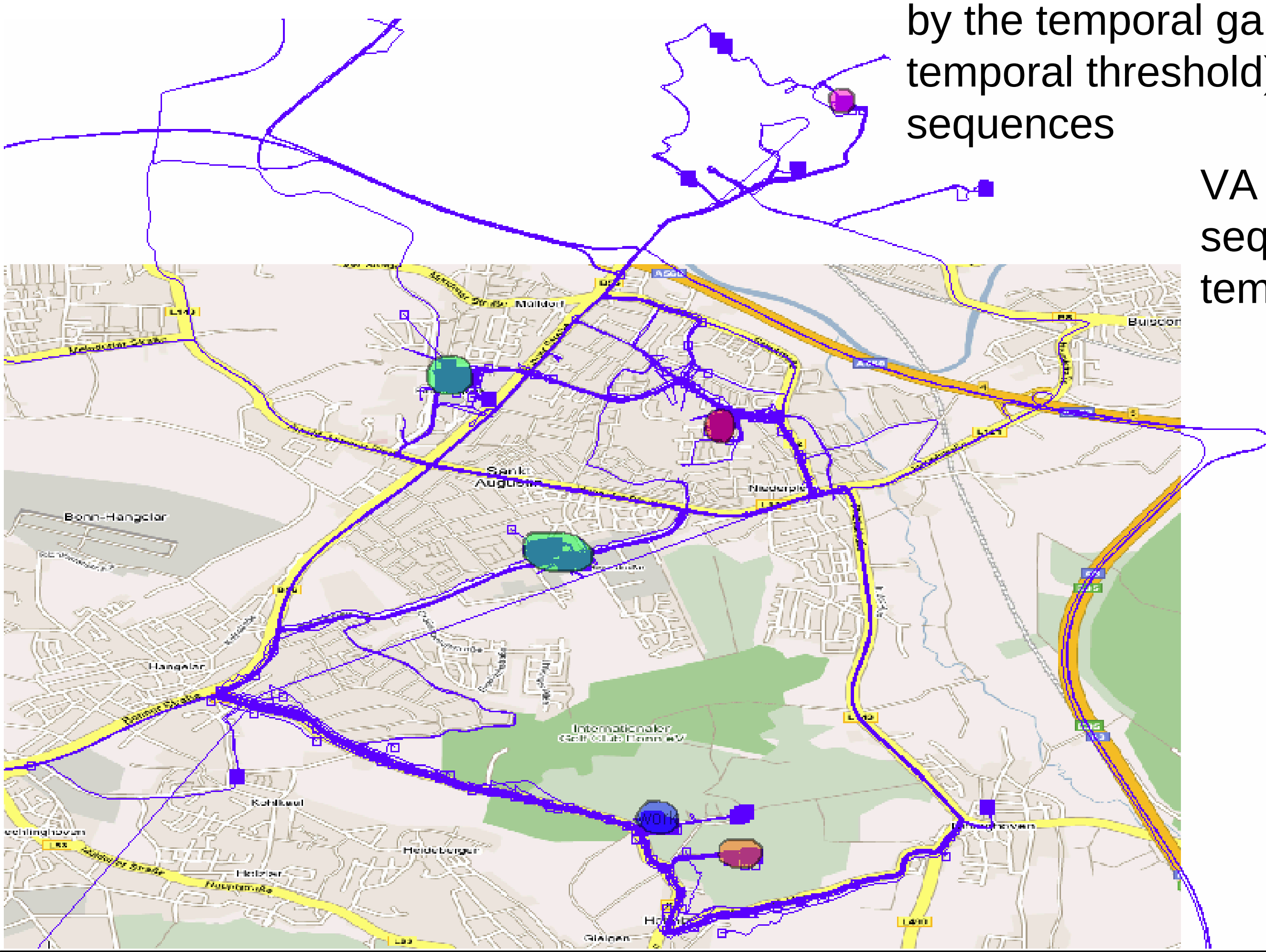




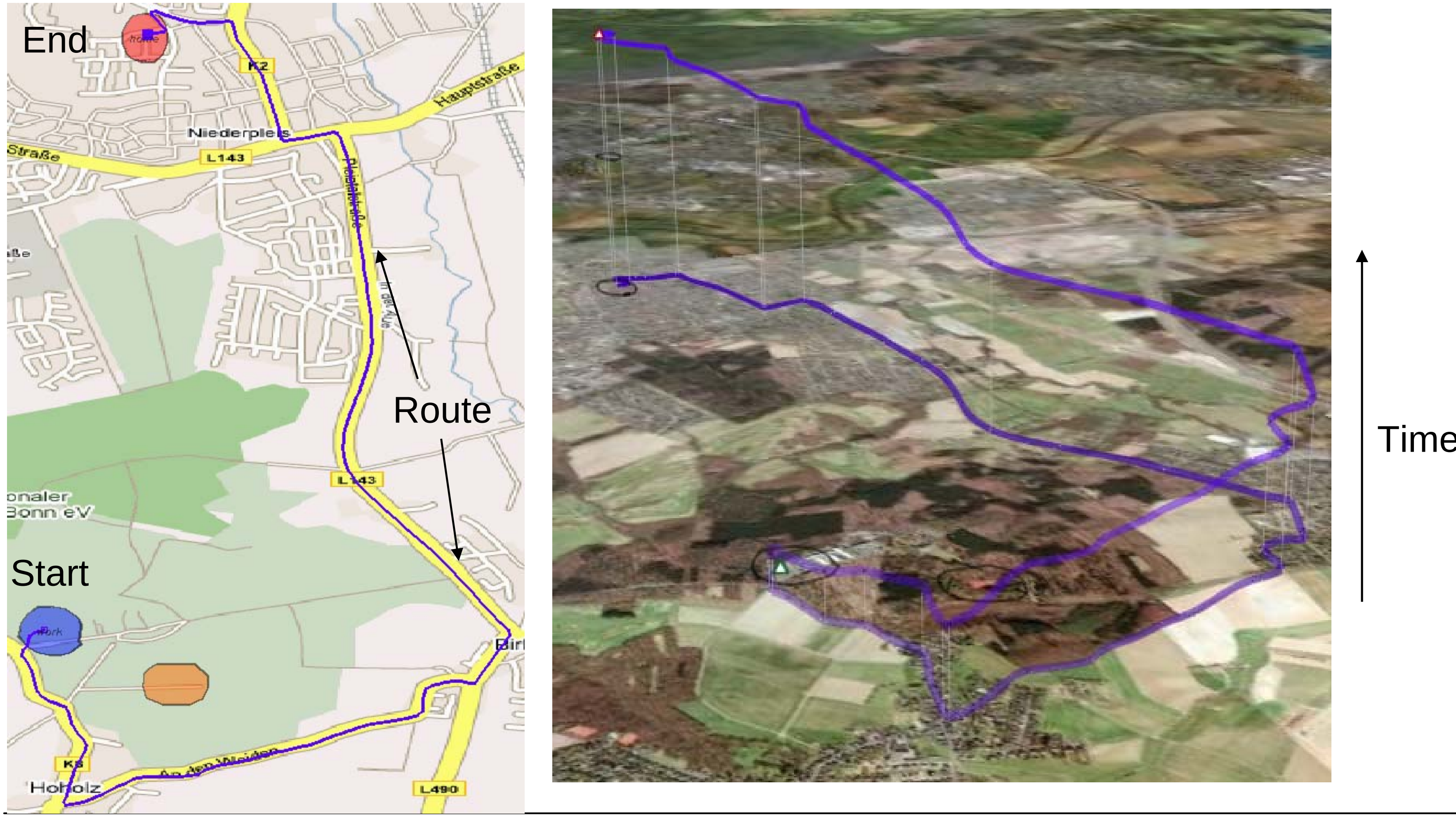
Extraction of trips (trajectories)

DB query: divide the points into sequences by the temporal gaps (DIFTIME >= chosen temporal threshold) and extract the sequences

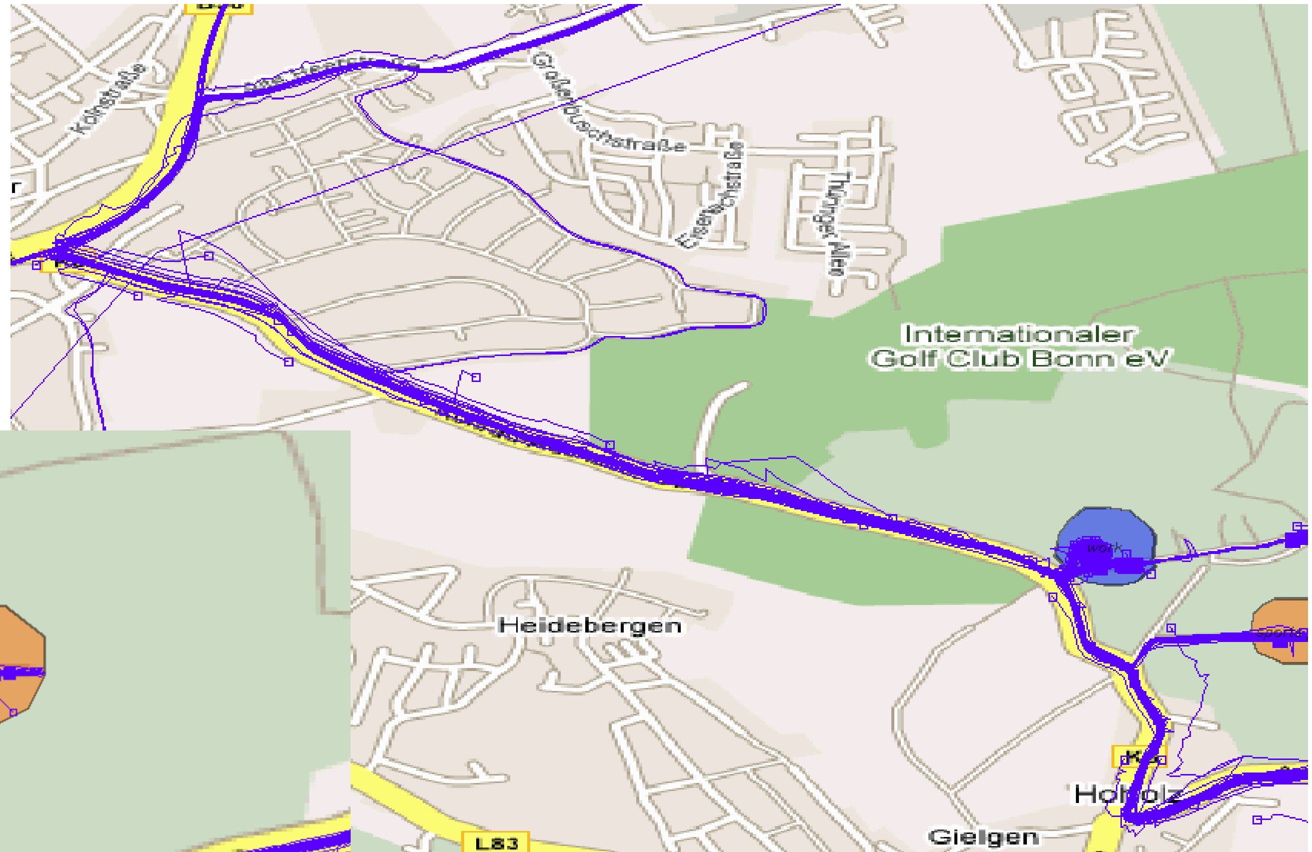
VA toolkit: represent the sequences as spatio-temporal lines (trajectories)



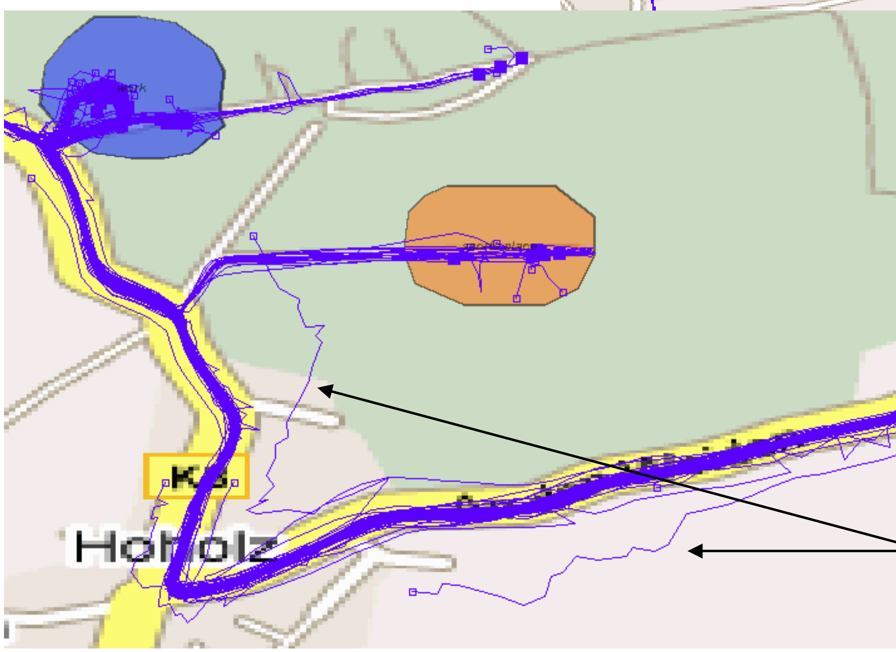
A trajectory in more detail



Problems with data quality…



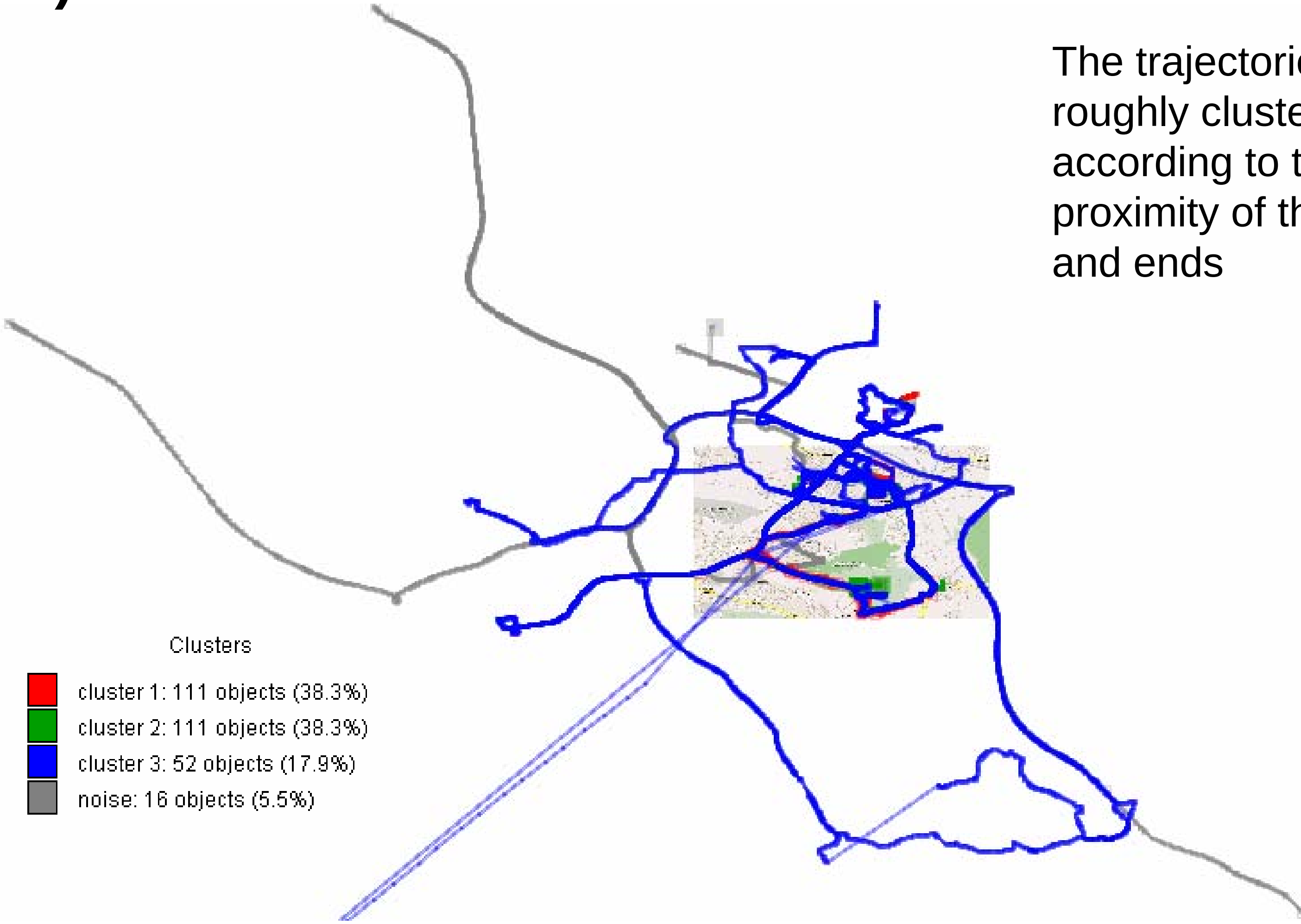
Many false starts;
the real starts are missing



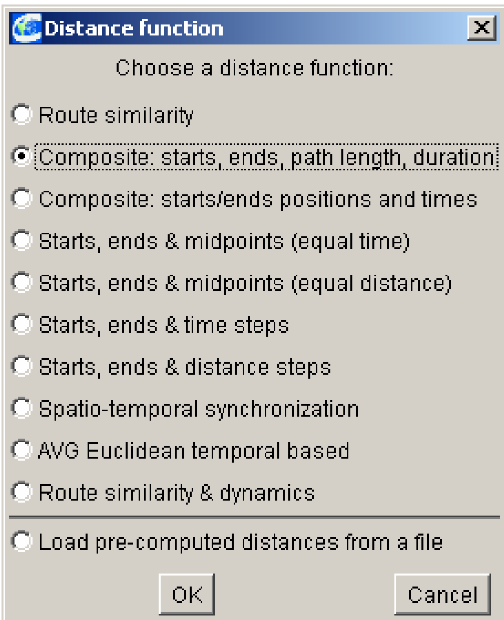
Positioning errors

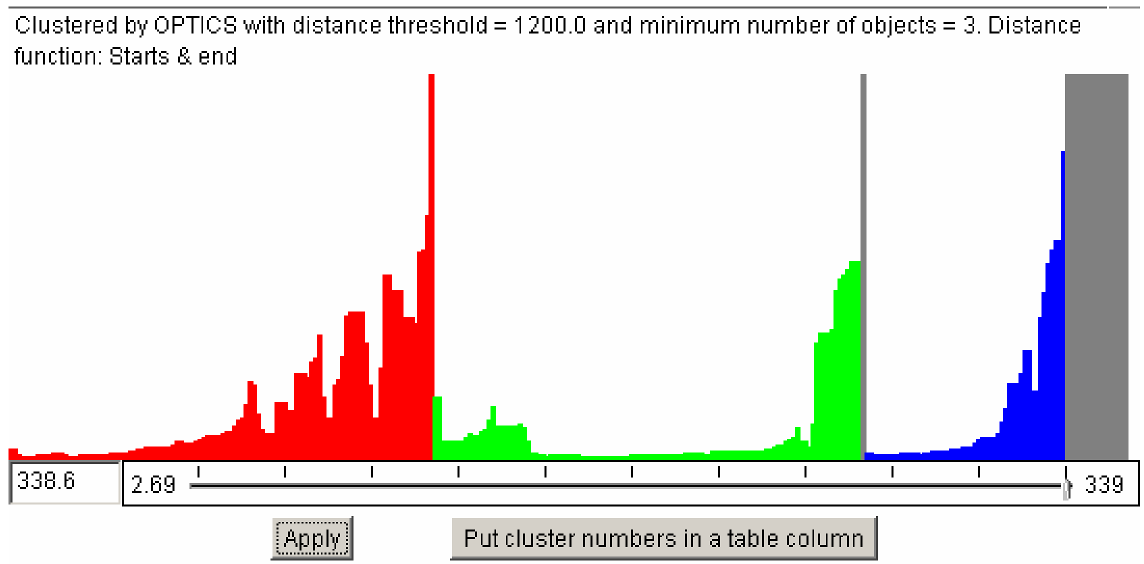
Looking for typical trips
1) From where and to where?

The trajectories are very roughly clustered according to the spatial proximity of their starts and ends



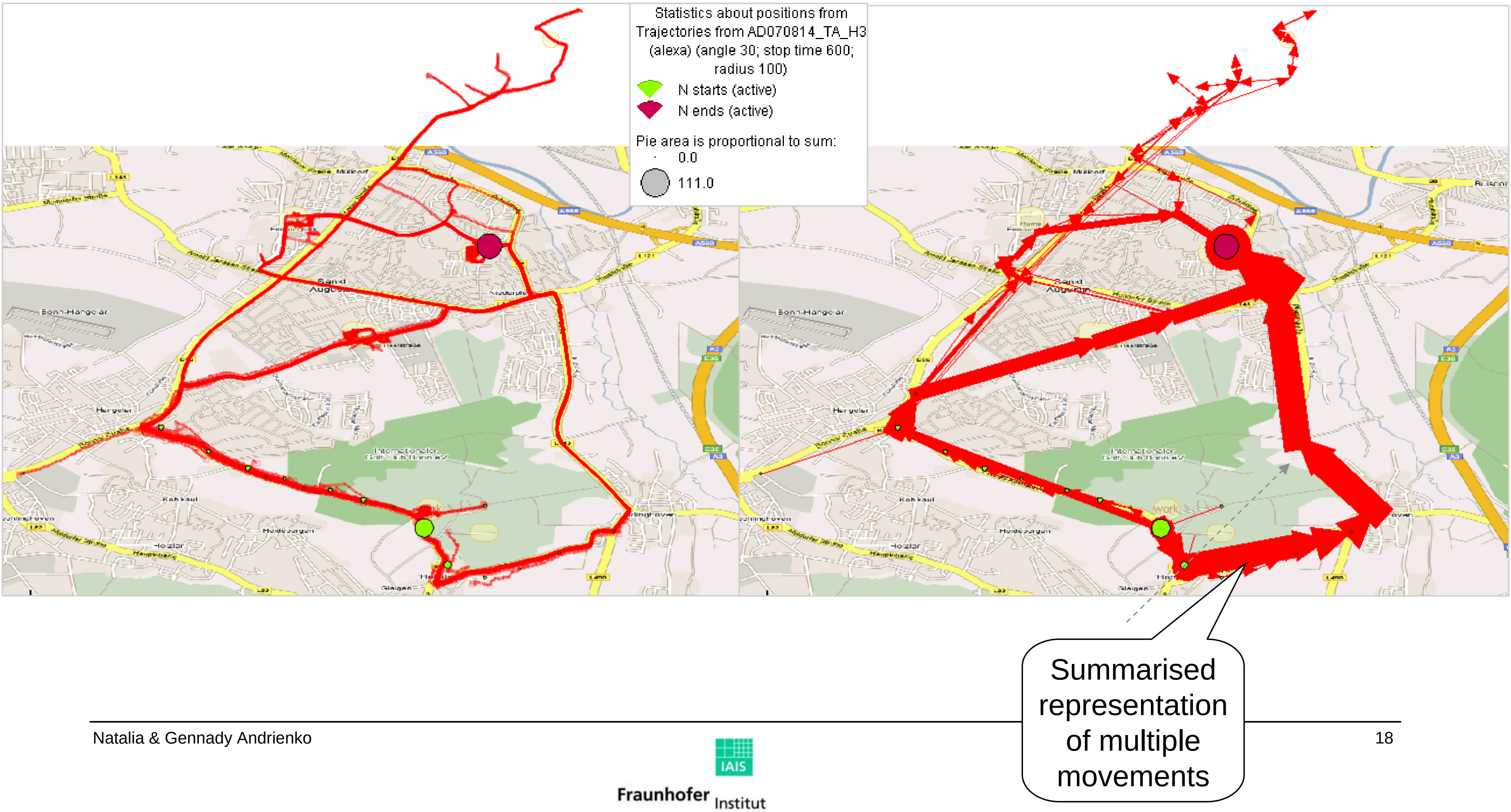
Interactive tool for spatial clustering



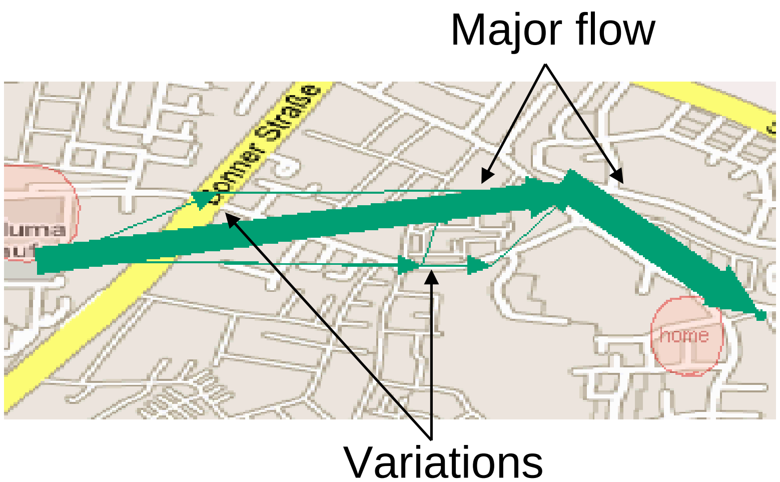


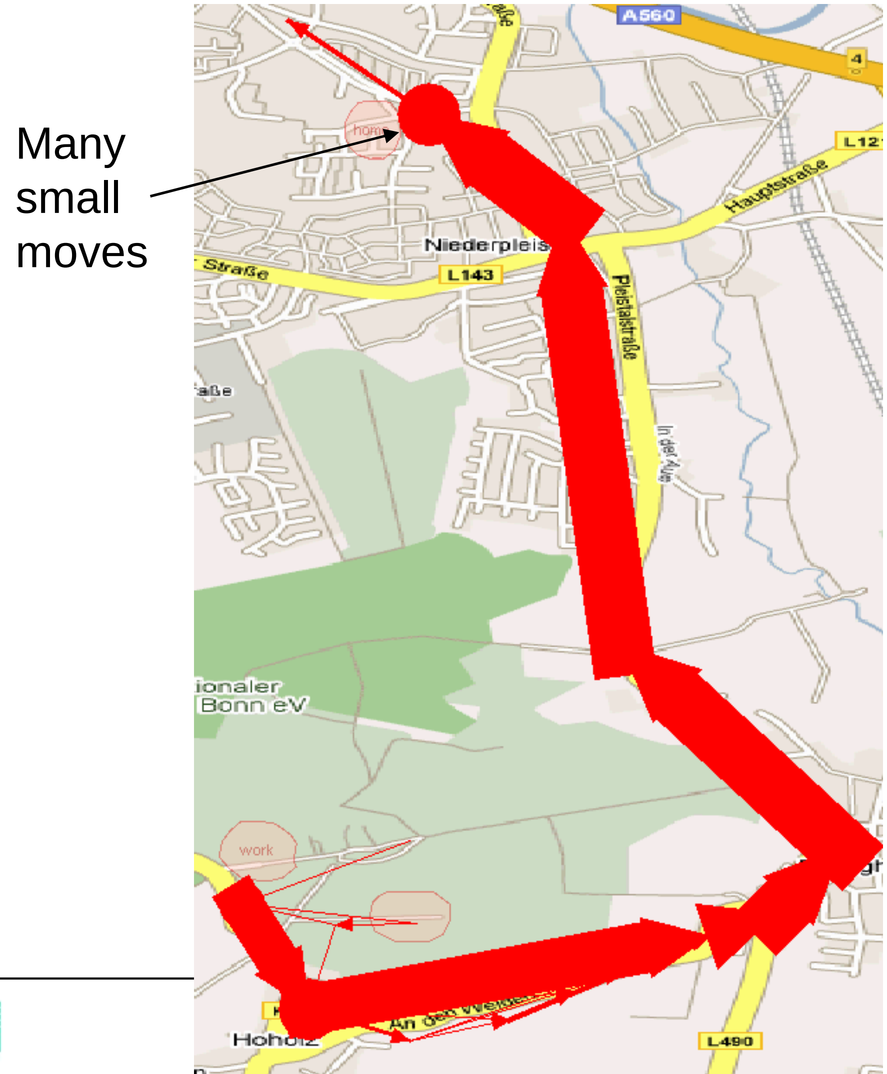
- spatial clustering algorithm OPTICS *(Ankerst, Breunig, Kriegel, and Sander 1999)*
- our implementation: cluster building is separated from distance and neighbourhood computation
- benefit: an analyst may try different distance measures
- there are several variants of distance measures designed specially for trajectories
- most of the measures can tolerate the specific errors in GPS-collected data

Cluster 1: from around the work to home (111 trips)



The summarised representation in more detail



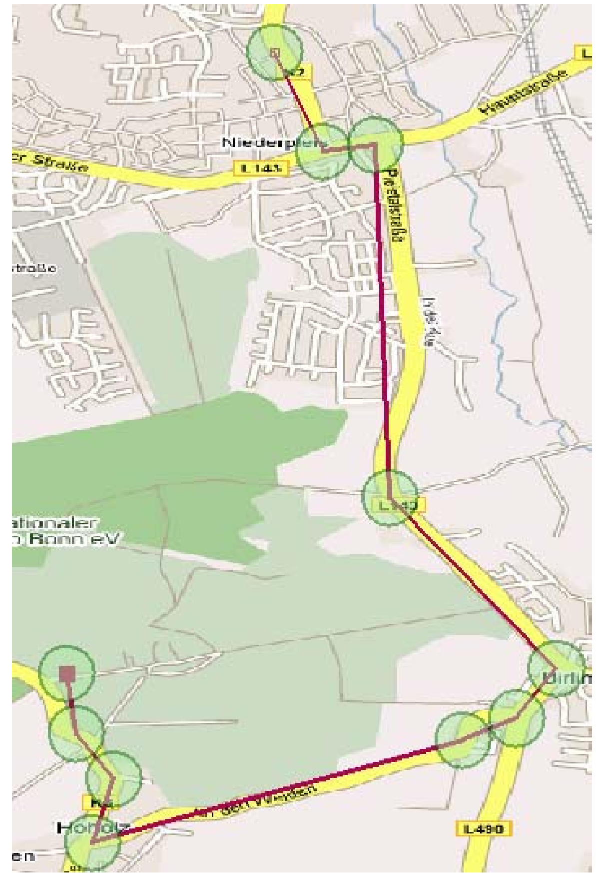


Summarisation of trajectories: how it is done (1)

Original trajectory

Simplified trajectory

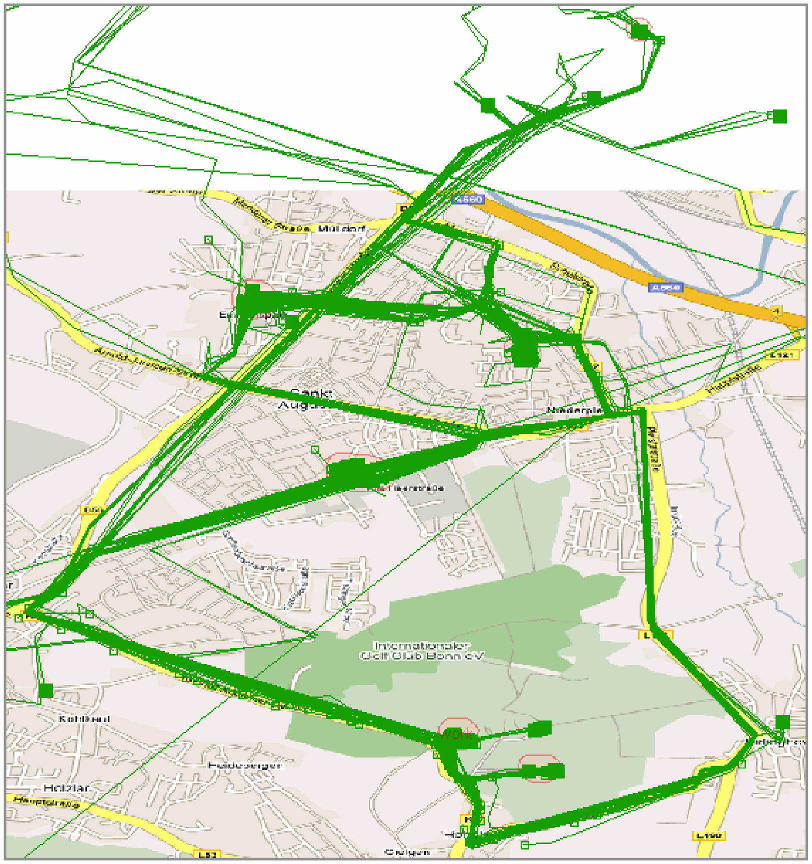
Generalised positions



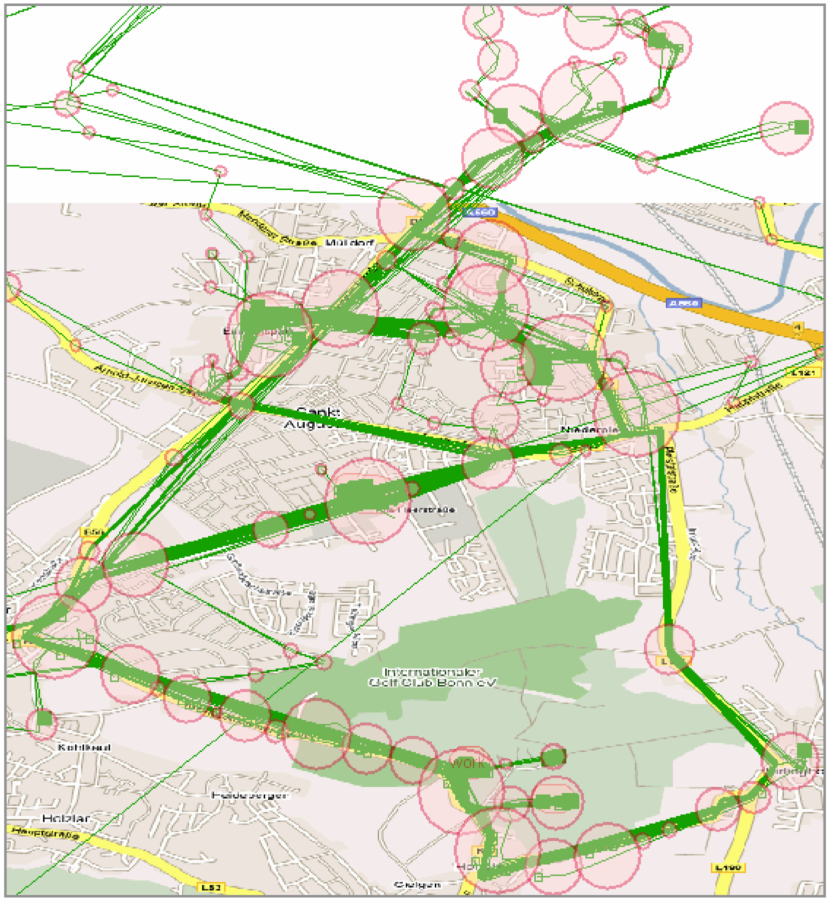
Minimum angle of direction change (degrees):	30
Minimum duration of a stop:	60
Minimum radius around a position:	100.0

← Specifies which positions to keep

Summarisation: how it is done (2)



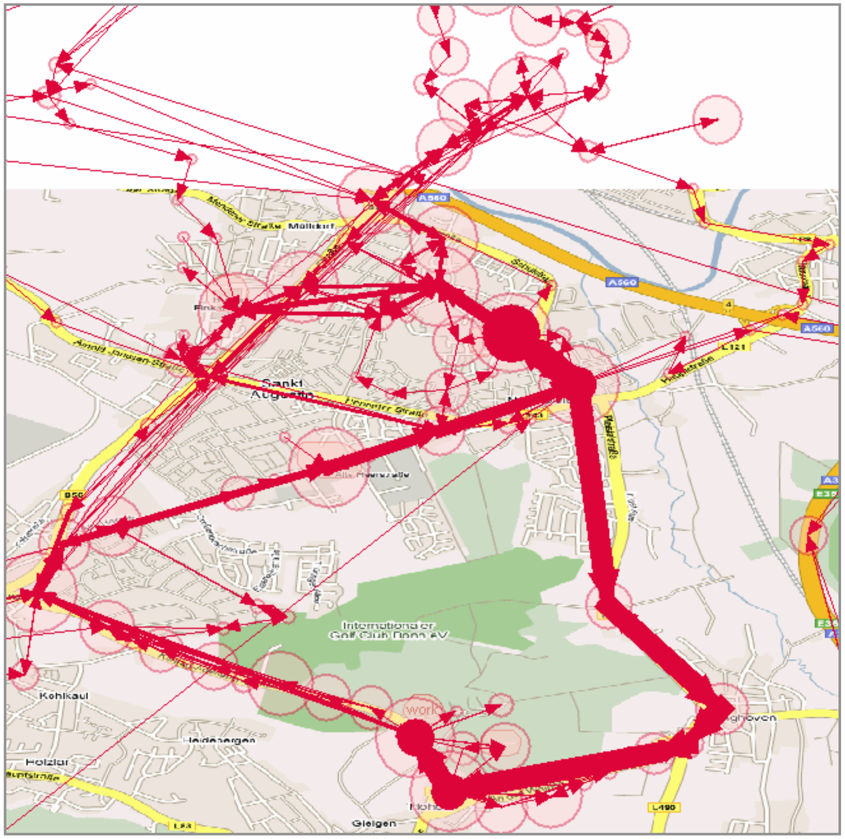
All trajectories simplified

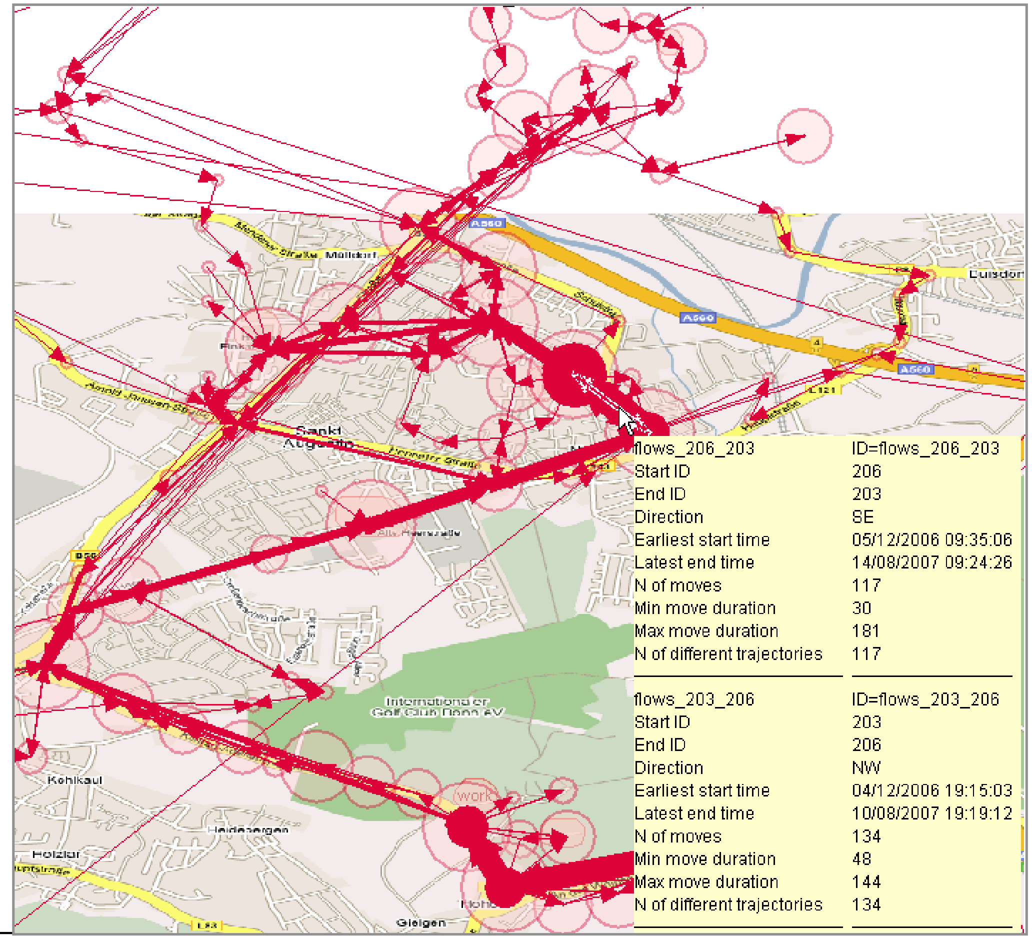


Generalised positions (circles, for simplicity) are built around spatially close positions of all trajectories

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

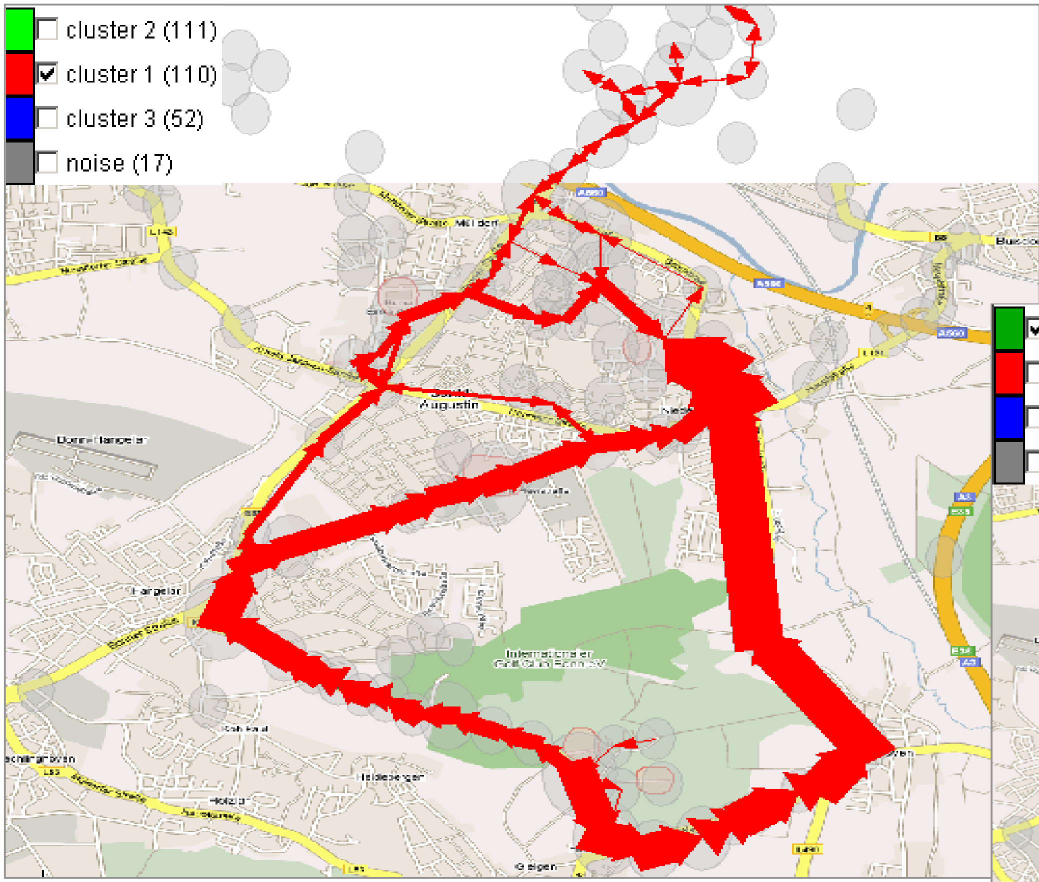
Summarisation: how it is done (3)



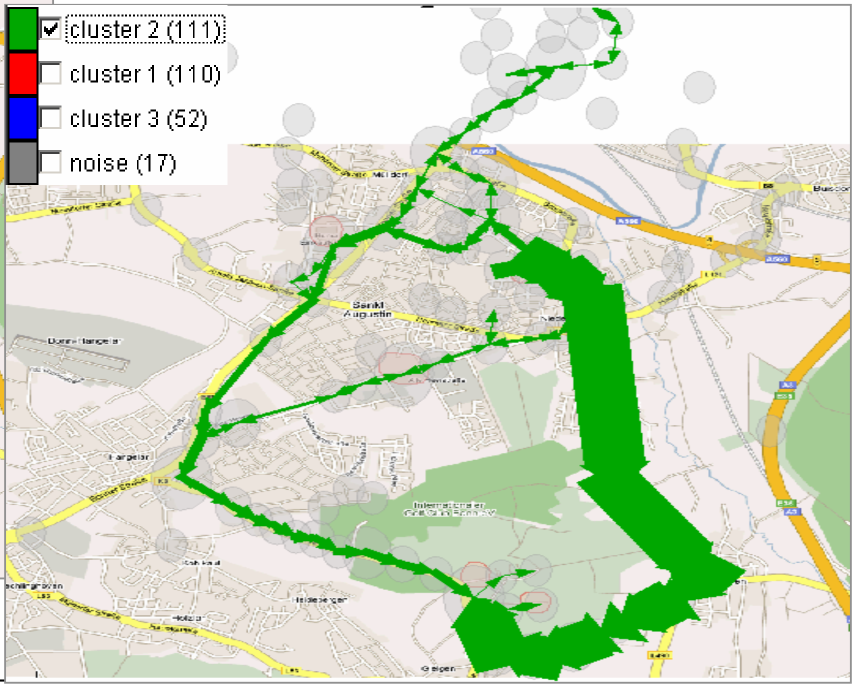


Numbers of moves between the circles are counted and represented by arrows of varying width

Summarisation: how it is done (4)

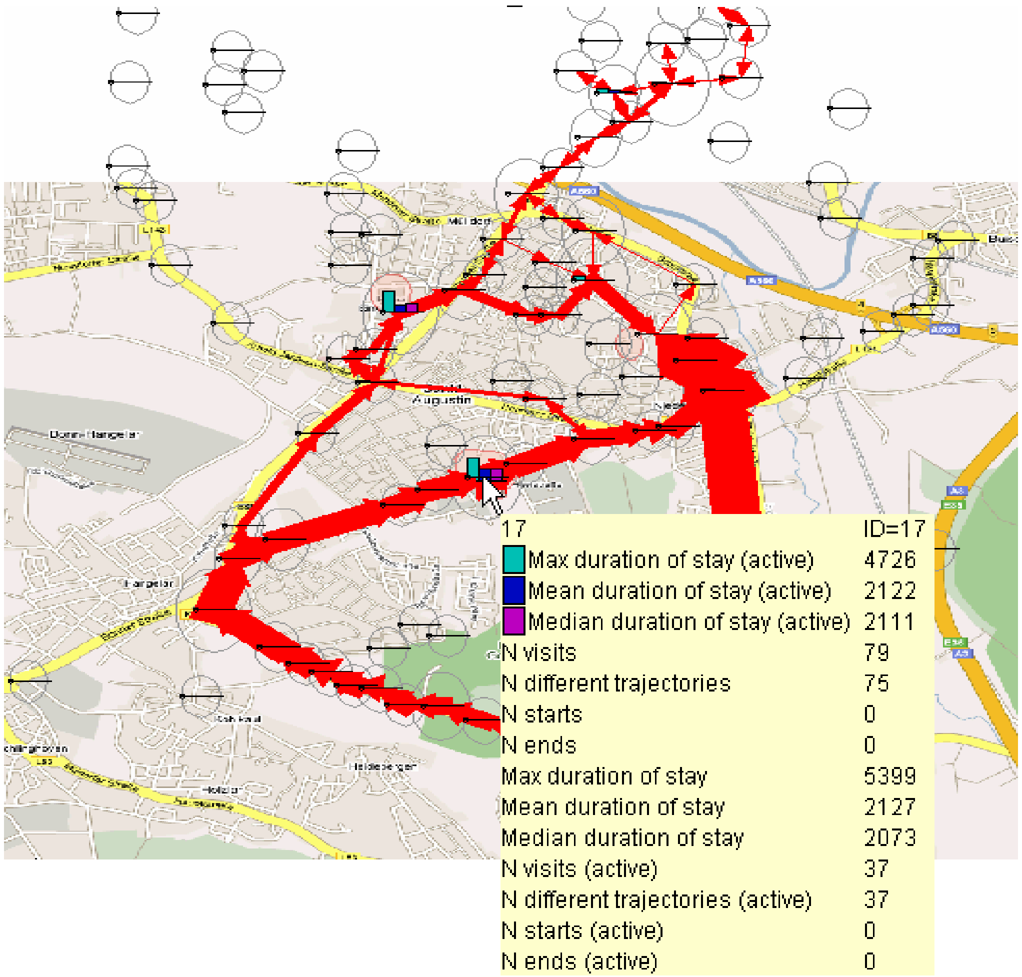


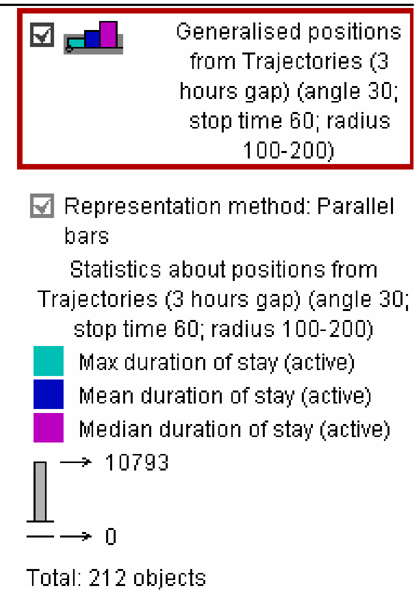
When a single cluster is selected *(or any other filter is applied to trajectories)*, the aggregate moves represent only those trajectories that are currently active



Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Summarisation: how it is done (5)

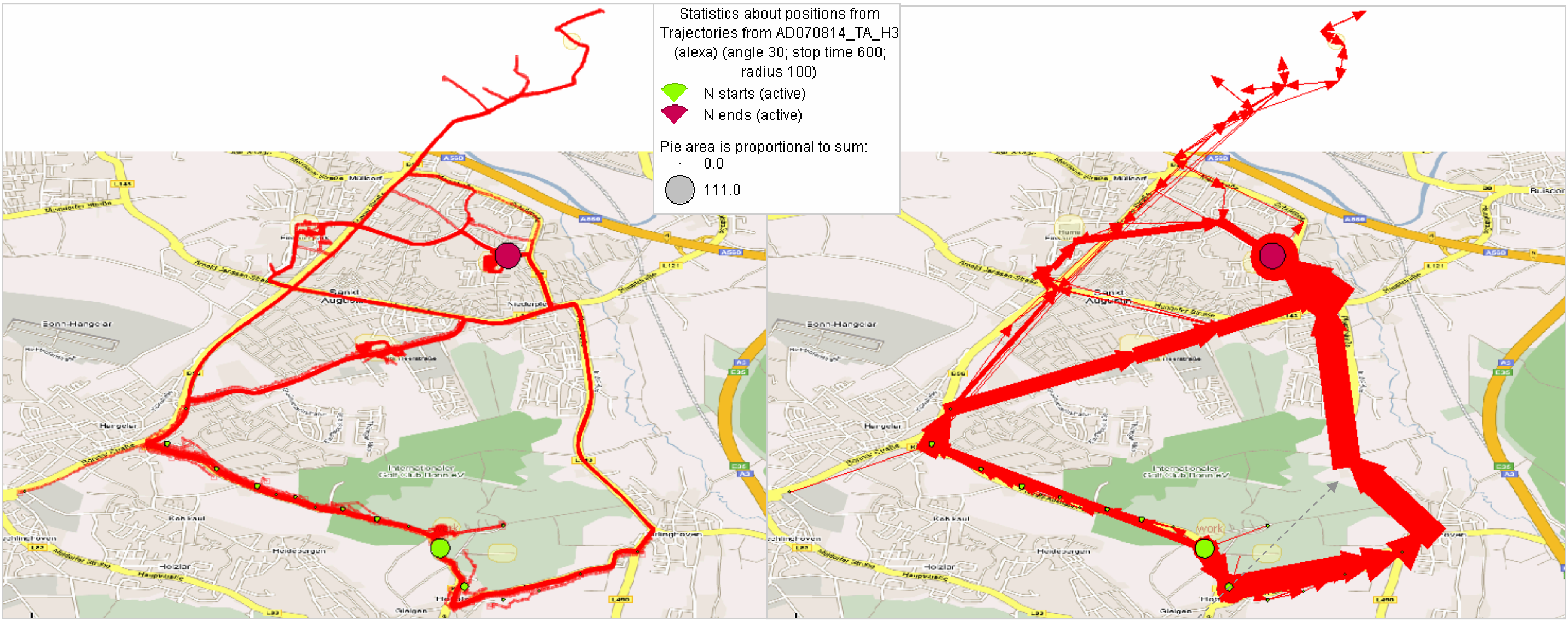




The movement data are also summarised by the generalised positions (circles).
When the selected subset of active trajectories changes, the corresponding attributes of the circles are re-computed (these are thus *dynamic attributes*).

Various visualisation techniques can be applied to both static and dynamic attributes of the generalised positions.

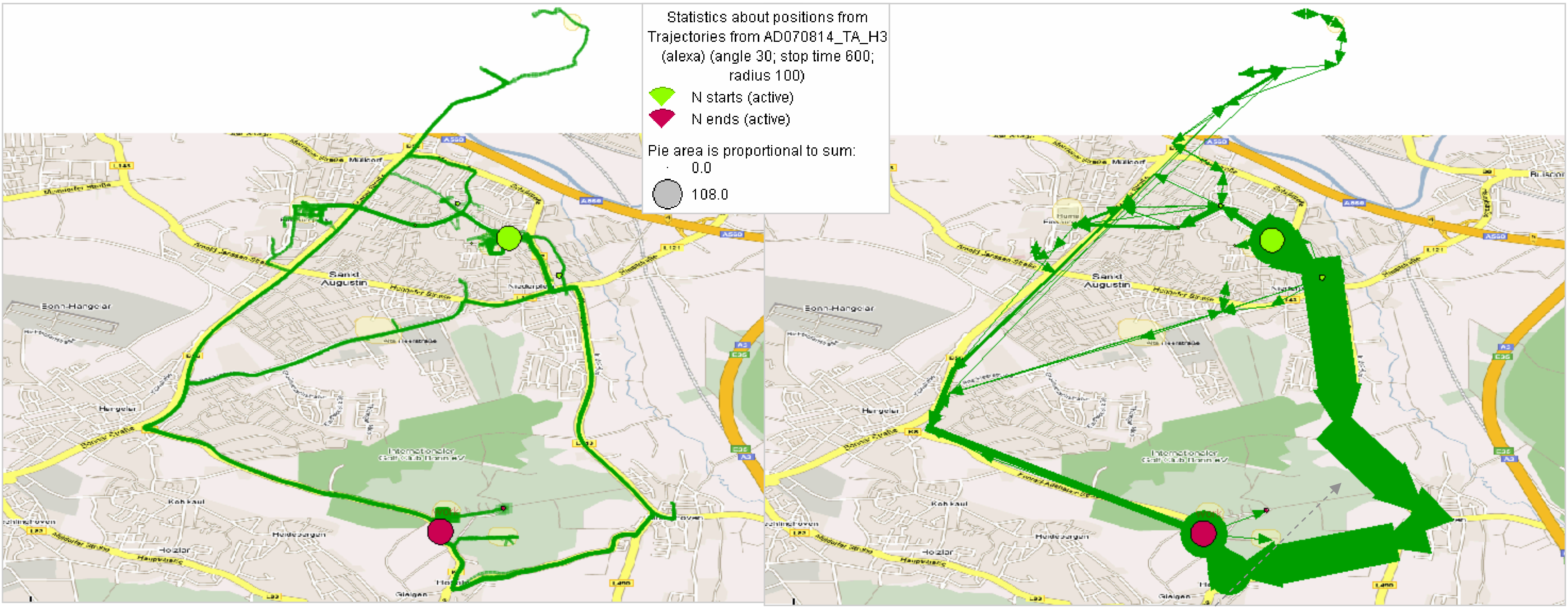
Cluster 1: from around the work to home (111 trips)



The "false starts" can be seen (small green circles)

Observation: the eastern route is chosen more often

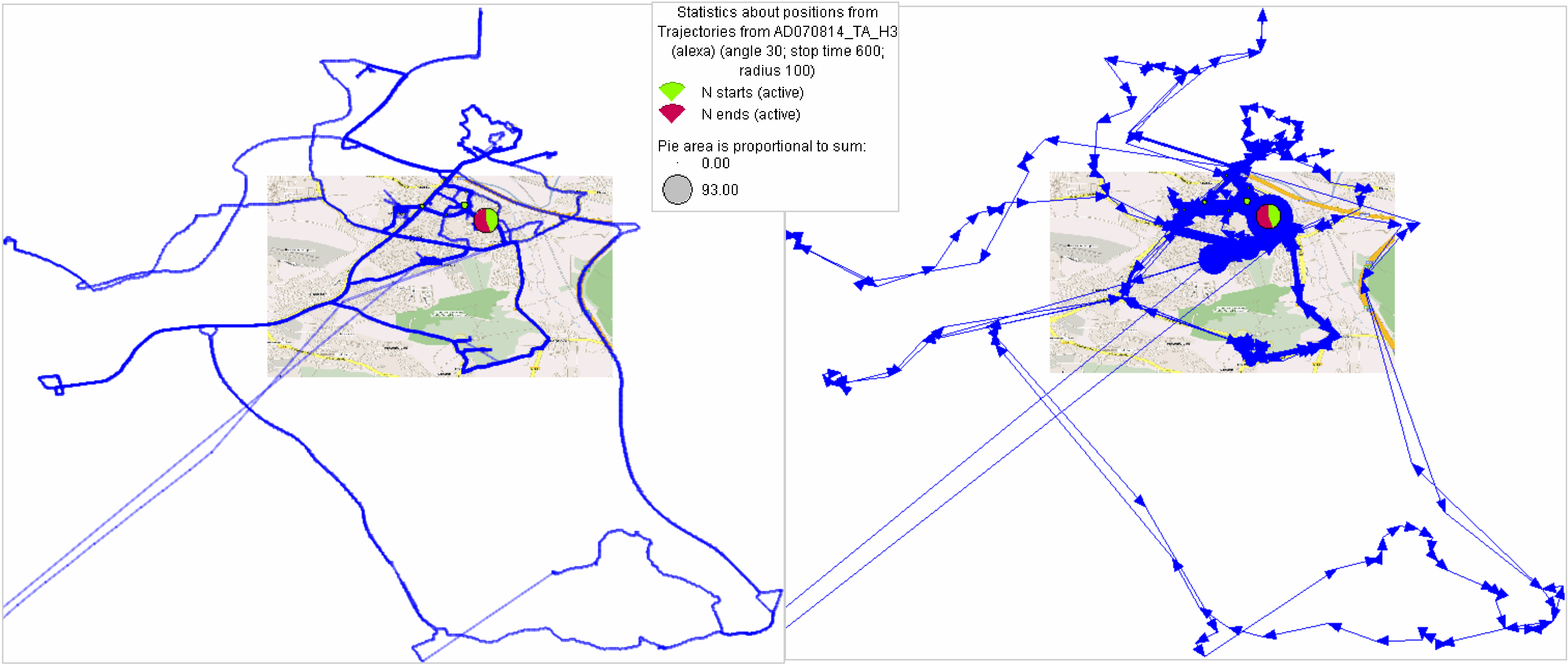
Cluster 2: from around home to around the work (111 trips)



Observation: the eastern route is chosen much more often

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Cluster 3: trips starting and ending around home (52 trips)

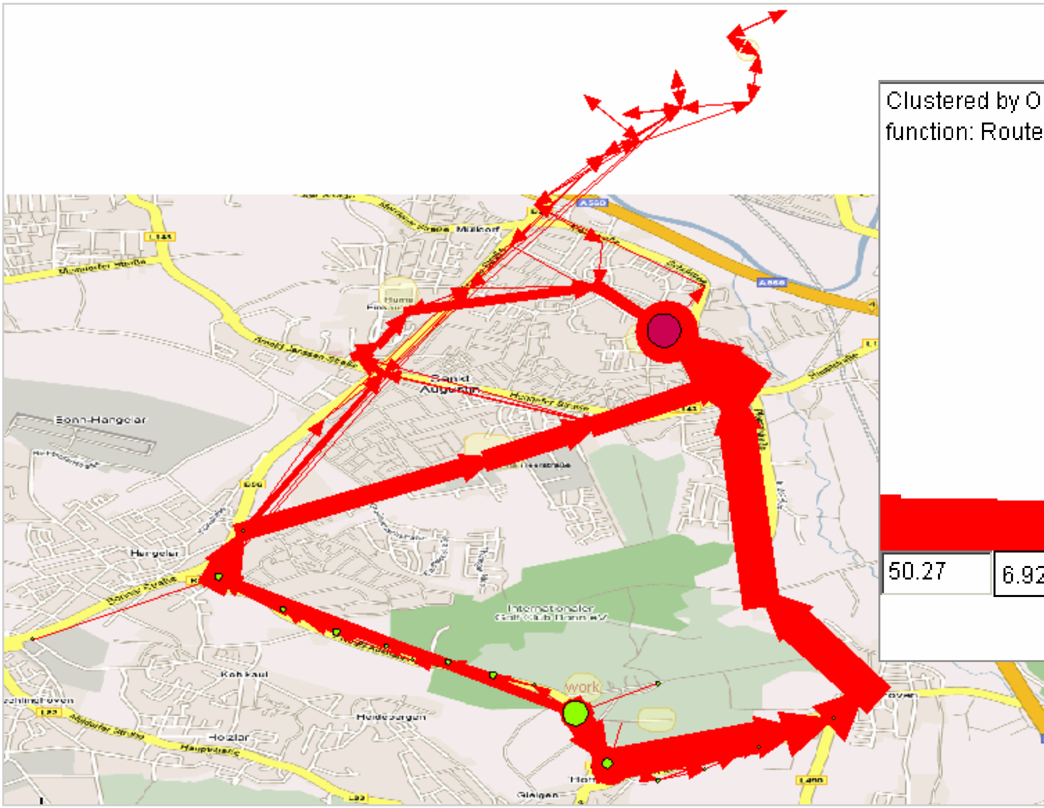


Observation: the distant trips are mostly occasional

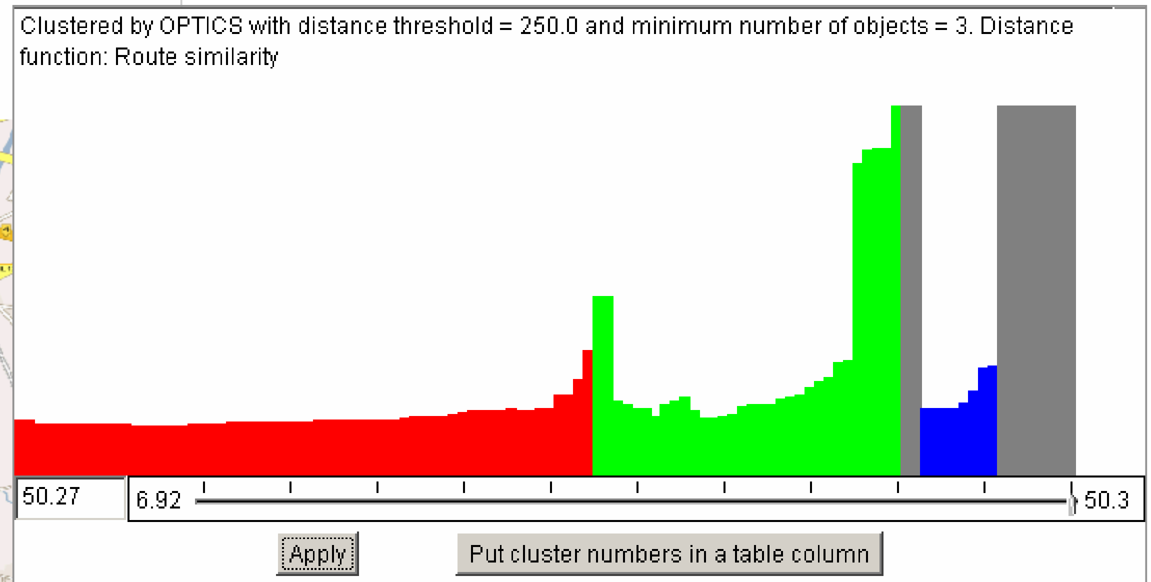
Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

Progressive clustering

1) Select one or a few clusters



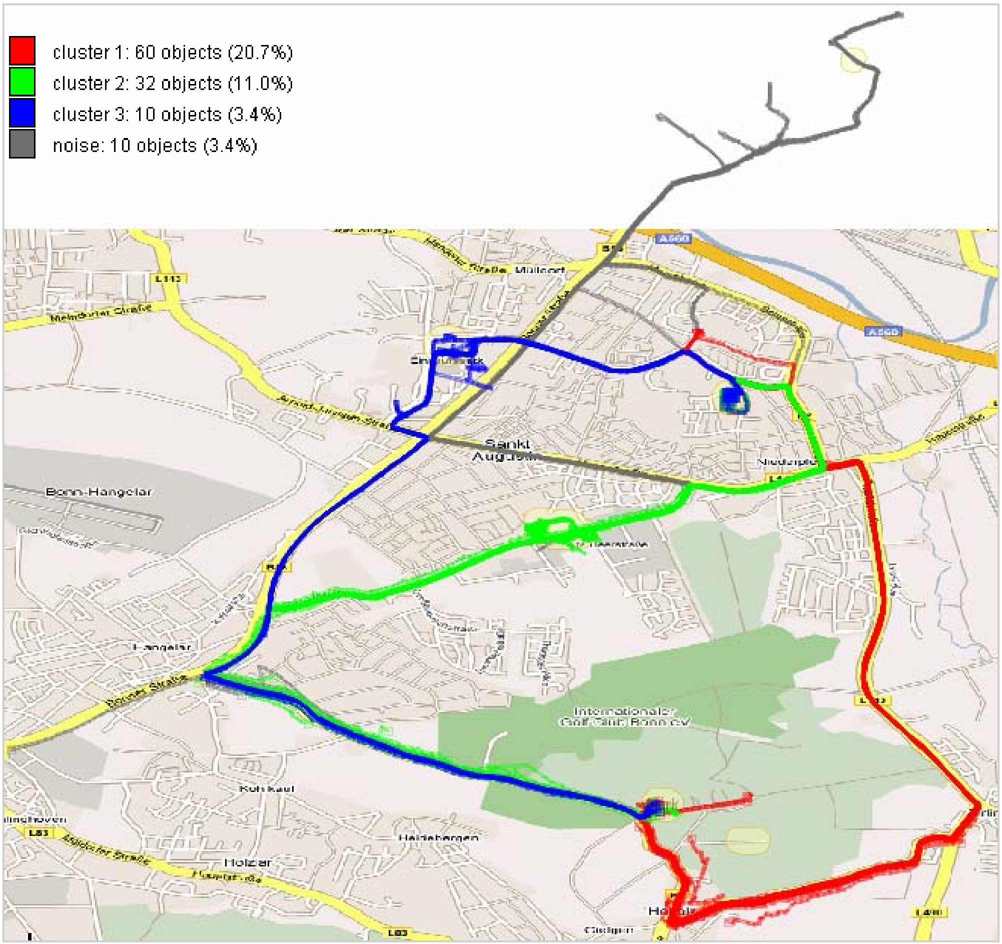
2) Apply another clustering to the selection; use a different distance function or different parameter settings



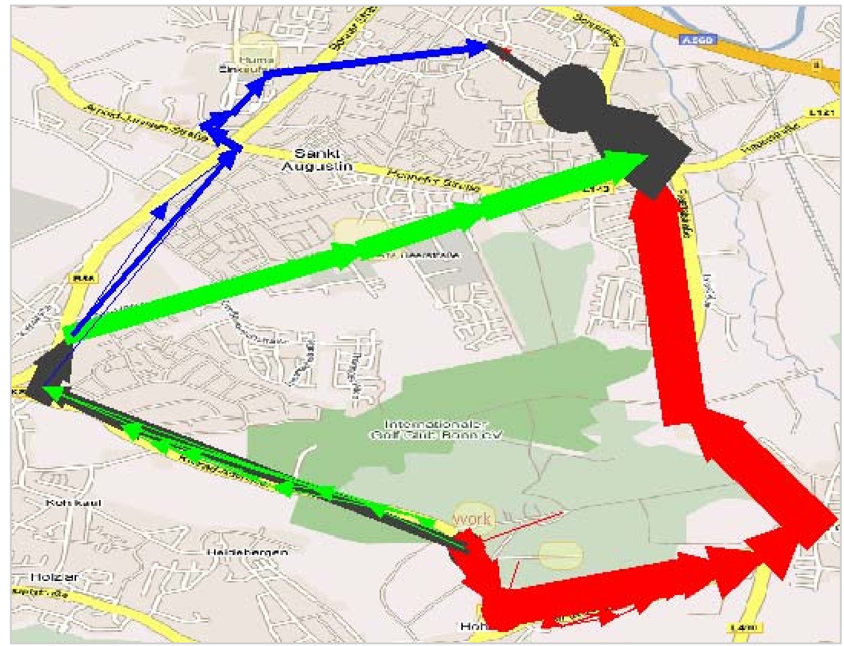
IAIS
Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Analysing the trips from work to home (111 trips)

Clustering according to route similarity applied to the cluster of trips from work to home



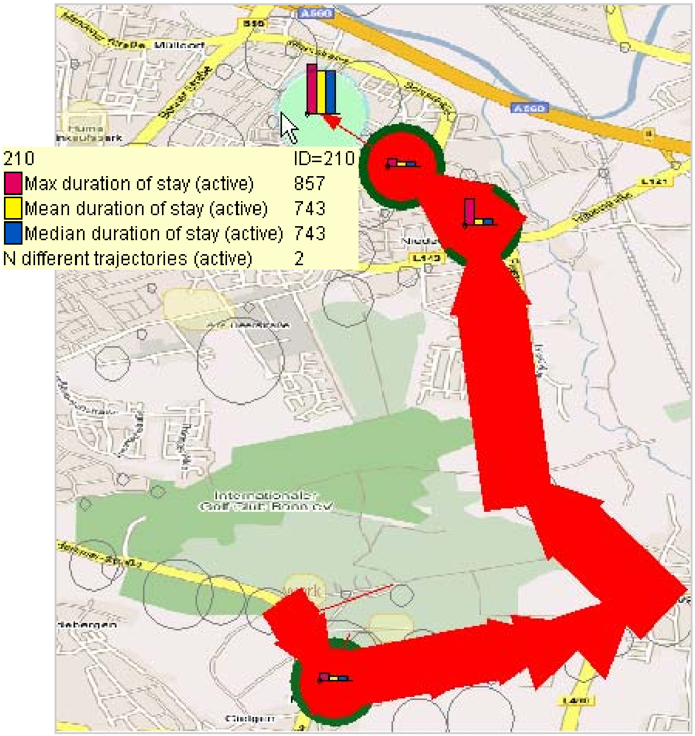
Summarised trajectories excluding the noise



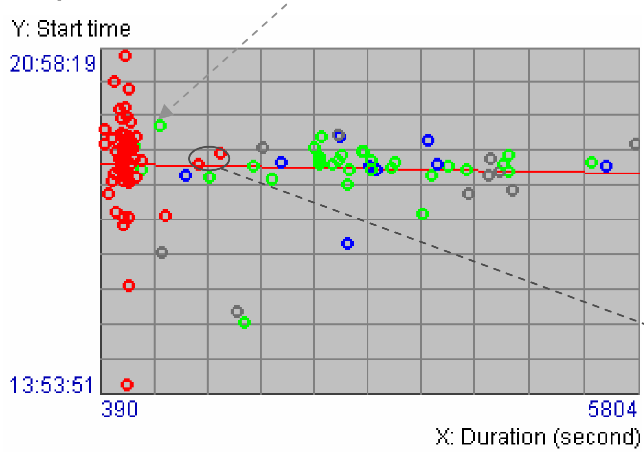
Hypothesis: one of the routes (red) is direct; the others are used for visiting certain places like shops.

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

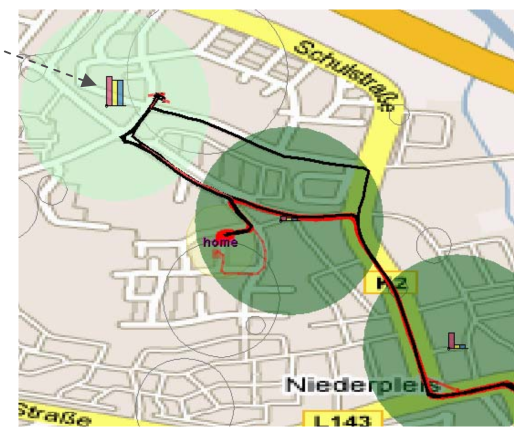
The “red” route (60 trips)



The “red” route takes less time than the others. Sometimes, the trips start quite late in the evening while the latest of the remaining trips starts at 19:22.



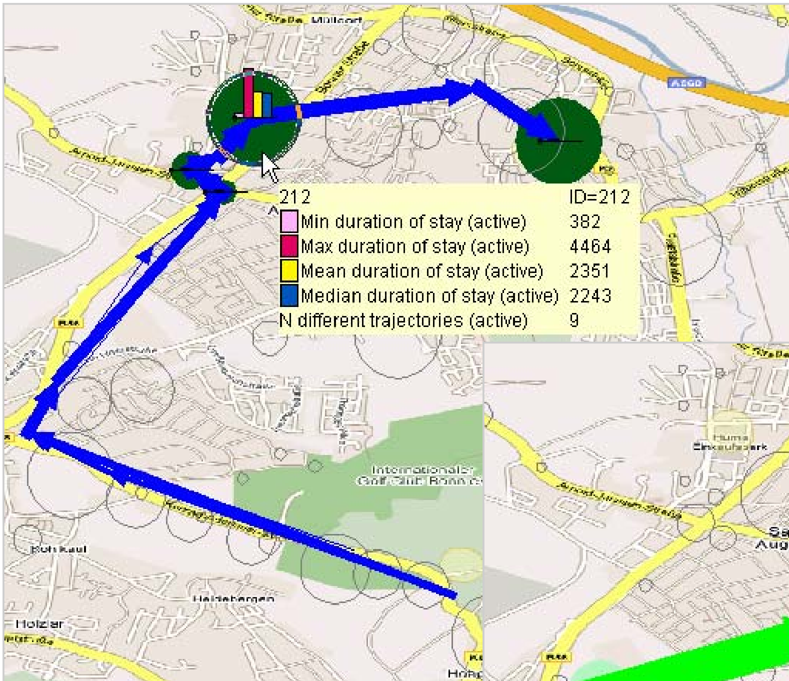
Here the time was spent

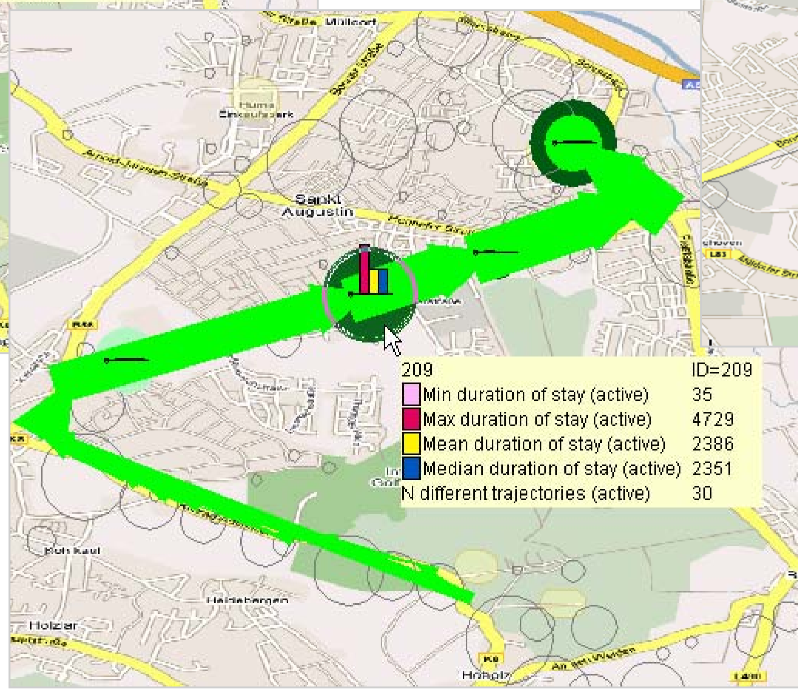


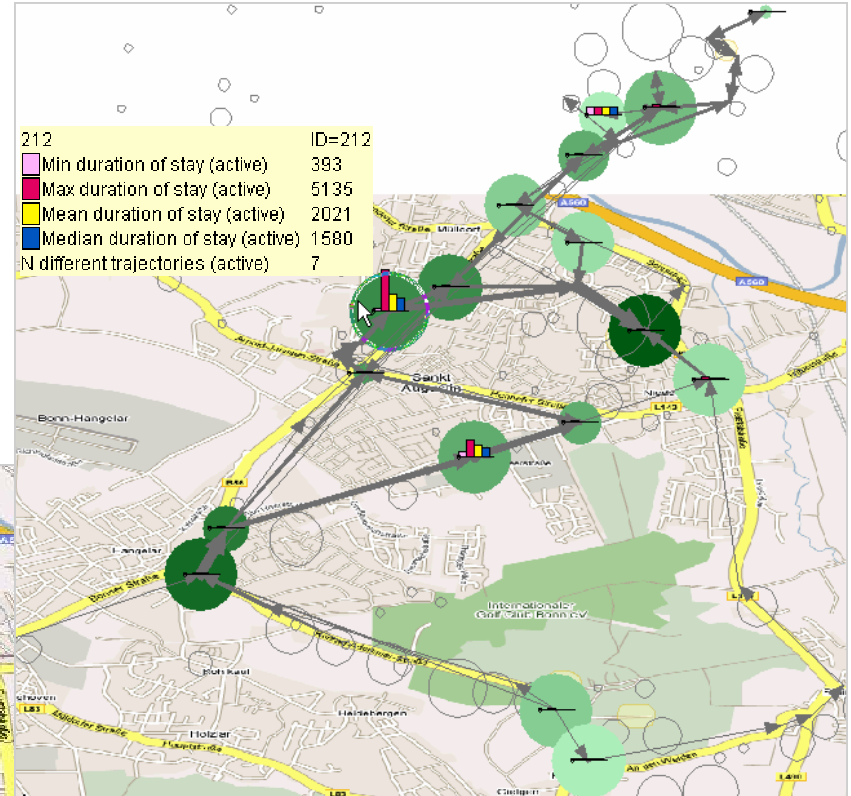
Stops are rare and rather short; the longest stops (just 2) were in a place not far from home on route prolongation

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

The other ways home







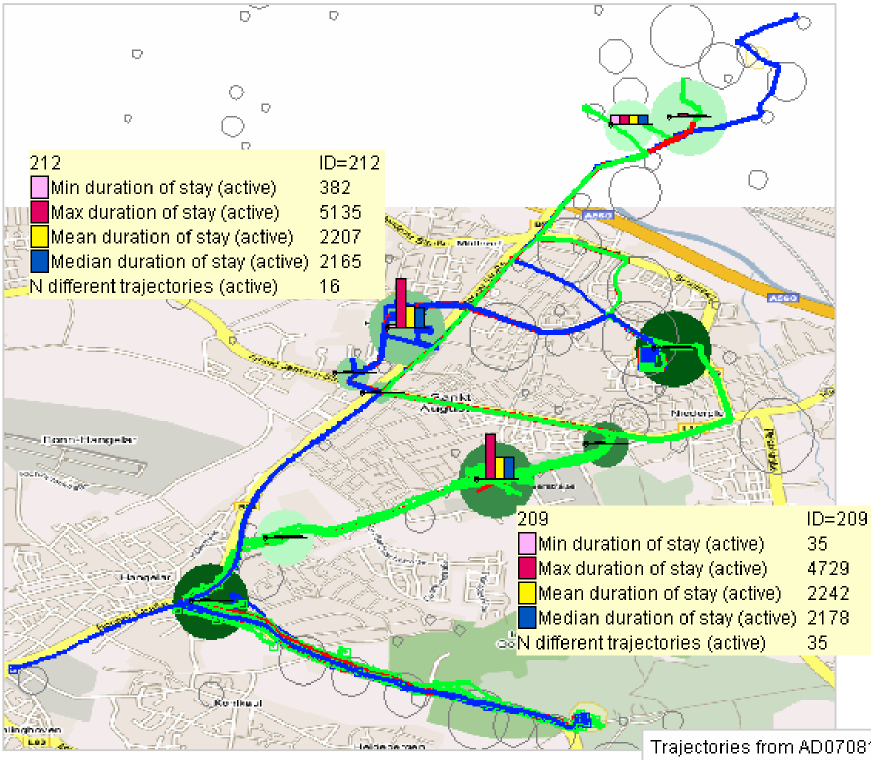
Our hypothesis is supported: the “red” route work-home is direct; on the other routes, the person makes intermediate stops.

On the way from the work to home, the person often visits the shopping centres. The route through shopping centre 2 (green) is preferred, but not always the person stops there (see min duration of stay).

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

The trips through the shopping centres

Is there any regularity in preferring one shopping centre to the other?



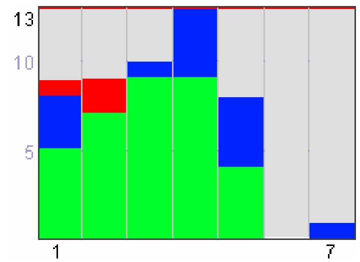
Trajectories from AD070814_TA_H3 (alexa): general data

Visits of shopping centres

- none: 61 objects (21.0%)
- 2: 34 objects (11.7%)
- 1: 13 objects (4.5%)
- both: 3 objects (1.0%)

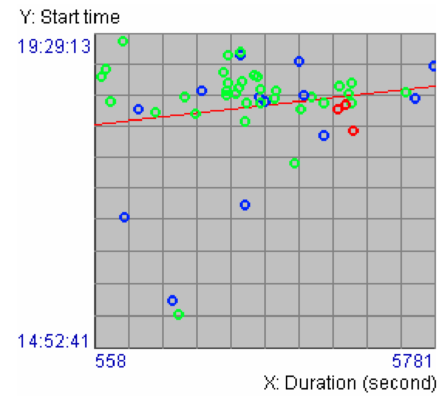
On the basis of the clustering results, we have interactively classified the trips according to the visits of the shopping centres.

Weekly distribution of the trips



Red: trips through both shopping centres

Start times and durations

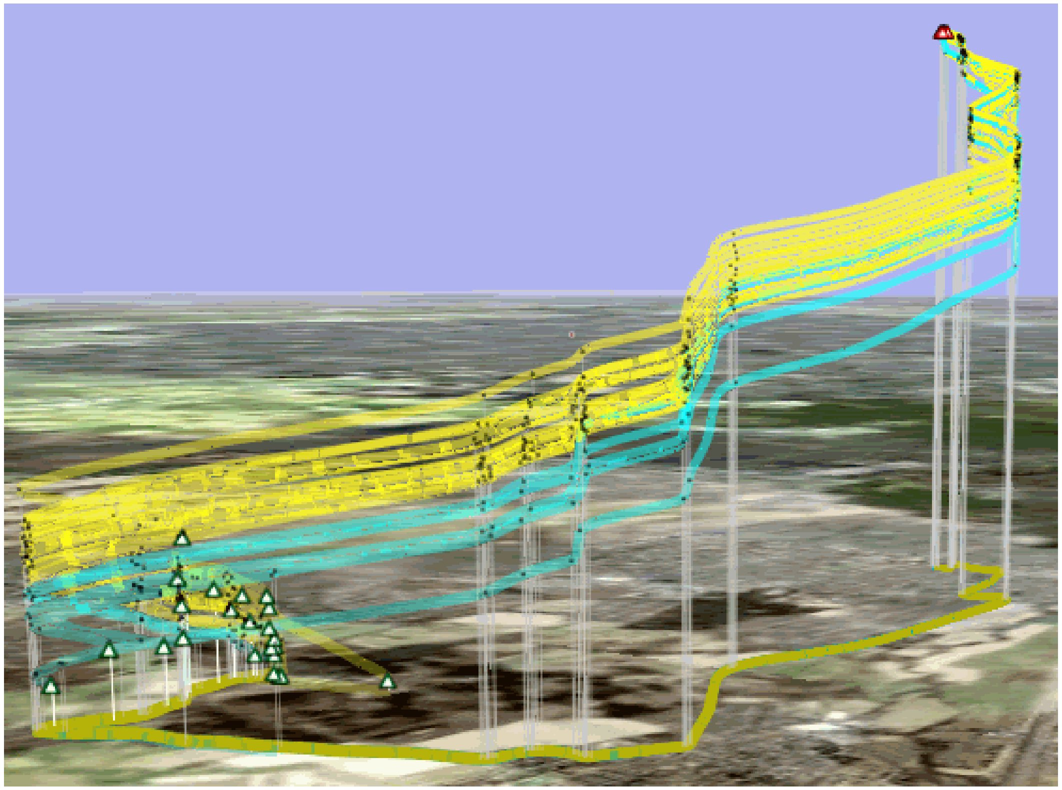


1) The trips through the shopping centres occur more often on Wednesdays and Thursdays.

2) Shopping centre 1 is usually visited on Mondays, Thursdays, and Fridays.

3) The starting times and durations of the trips through shopping centres 1 and 2 do not significantly differ.

The use of 3D view (Google Earth) to analyse trajectories

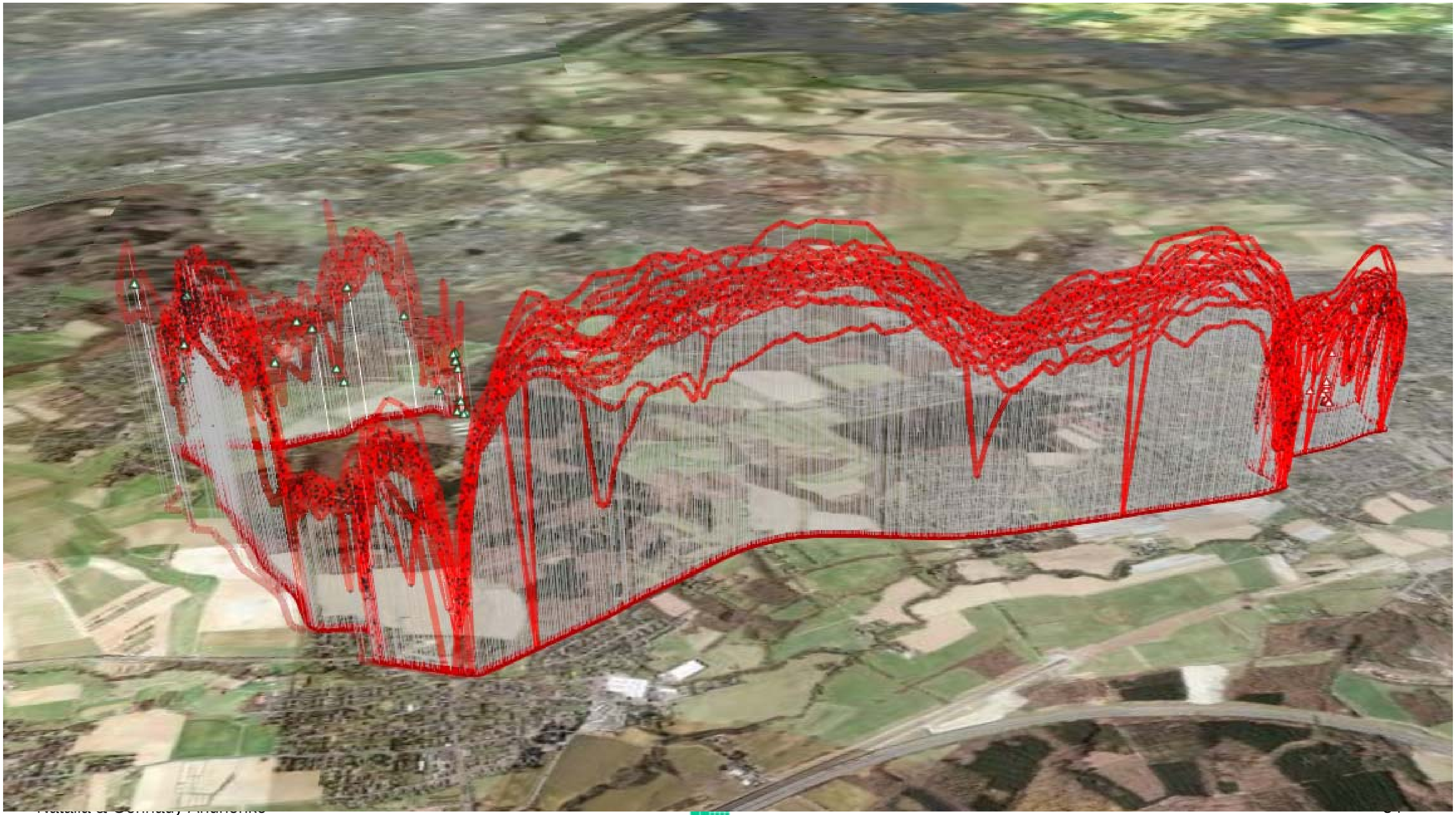


Two clusters of trips from home to work by the direct route differing in dynamics
Yellow: faster trips; cyan blue: slower trips

The clusters obtained using the distance function “route + dynamics”

IAIS
Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

Comparison of speed dynamics in 3D view (Google Earth)



What we have learned about the car owner:

- The places where the person lives, works, and shops
- The typical routes from home to work, from work to home, to the shopping areas
- The places where the person frequently stops on the way from work to home
- The durations of the stops, times spent for visiting the shopping areas
- The times of the trips and of the stops
- How the chosen routes are distributed over the days of the week

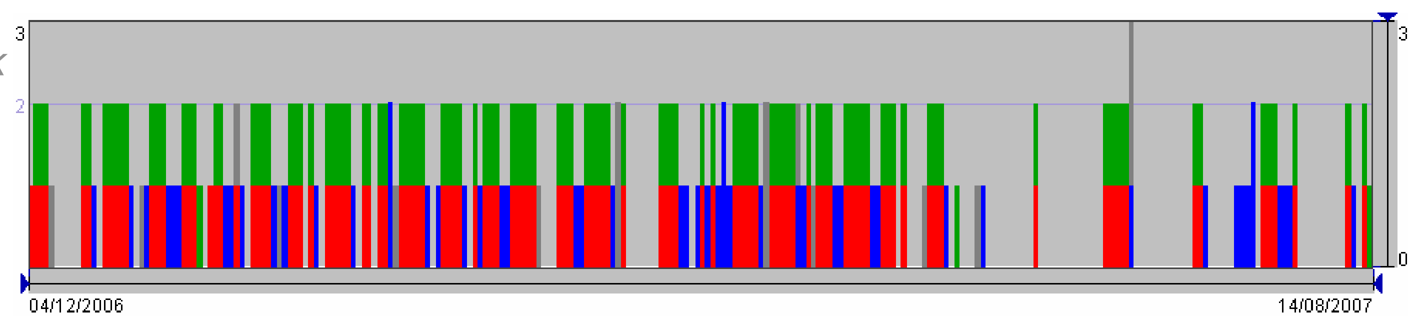
The tools we used:

- Database processing
- Interactive clustering
- Computational generalisation and summarisation
- Interactive displays
- Dynamic filtering

What we also have learned but haven’t noted explicitly:

- The car owner lives in a small town (found using Google Earth and Google Maps)
- The person has a flexible work schedule (concluded from the variation of the times of the trips to and from the work)
- The person has no small children (concluded from the times of the trips from the work to home, which are often quite late)
- The person is often away or sick (judging from the distribution of the trips over the time period, especially in the summer)

green: home to work
red: work to home
blue: home – home



Serious privacy problems ⇒ Individual data must not be accessible!

IAIS
Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

Reconstruction of the analytical process

raw data (~90K records <x,y,t>)

Database processing

enhanced data (~90K records <x,y,t,dt,dx,dy, speed…>)

Interactive UI + Database processing

Locations of stops with variable *t*-resolution

Cluster analysis + visualization + Google Earth (additional data!)

Clusters of locations, their interpretation

Interactive UI + Database processing

Extracted trips (~500)

Interactive UI + geocomputations

aggregated trips

Cluster analysis + visualization and interactive analysis tools + Google Earth

Typical trips, their interpretation and analysis

Summary

- Movement data: time-referenced positions of some moving object(s) { + values of time-dependent attributes, e.g. characteristics of the movement }
- Analysis of the movement behaviour of an individual object:
  - Significant places
  - Typical trips and their characteristics: start time, duration, origin and destination, intermediate stops, speed dynamics
  - Atypical trips, e.g. unusually long trips home
- Analytical instruments:
  - Database queries: places of stops, sequences of positions
  - Spatial clustering: typical places of stops (significant places); typical trips by origin and destination and by routes
  - Progressive clustering of trajectories: gradual refinement of our knowledge about the movement behaviour
  - Summarisation of trajectories: generalised places, aggregate moves, dynamic (filter-sensitive) aggregation of moves between places and visits of places

See also

- Gennady Andrienko, Natalia Andrienko, Stefan Wrobel
  Visual Analytics Tools for Analysis of Movement Data
  ACM SIGKDD Explorations,
  2007, v.**9** (2), pp.38-46

Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme